# Transcriptomic analysis of *Clostridioides difficile* grown in a biologically relevant growth medium

Nicola Moore BSc (Hons), PgCert, MRes

Faculty of Life and Health Sciences of Ulster University

A thesis submitted for the degree of

Doctor of Philosophy

February, 2021

I confirm that the word count of this thesis is less than 100,000 words

*For Eva and Noah*

"*I was taught that the way of progress is neither swift nor easy*."

- Marie Curie

# TABLE OF CONTENTS

## ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Dr Nigel Ternan for his help and support since I first began my undergraduate placement many years ago. He has provided me with endless opportunities (and with seemingly endless patience) and for that I will always be grateful.

I would also like to extend my gratitude to Dr Chris Gill, Professor Geoff McMullan, Professor James Dooley and Dr Barry O'Hagan for all of their advice and expertise along the way. I would especially like to thank Dr Michelle Rudden for passing on her knowledge of all things qPCR. To everyone who has helped me with my lab work including those who provided technical support, thank you. Thank you to everyone from the lab, both past and present, who have offered their support and much-needed kindness during the toughest times.

I will be eternally grateful to all of my family especially my Mum and Dad for their encouragement and unwavering belief in me, thank you. Finally, I would like to sincerely thank my husband Aaron and our children Eva and Noah, without whose love and understanding I would have never made it this far.

## ABSTRACT

*Clostridioides difficile* Infection (CDI) is a known healthcare acquired disease and significant source of antibiotic-associated diarrhoea (AAD), reported to be responsible for up to a third of recorded instances worldwide. Most of the research conducted into how *C. difficile* causes disease is carried out in artificial laboratory media, limiting insight into how this pathogen responds to the environment in which it causes disease i.e. the large intestine. The overall aim of this study was to construct a faecal water model to study the transcriptome of *C. difficile* 630 in order to characterise the molecular mechanisms underpinning pathogenesis within the organism. We hypothesised that understanding how *C. difficile* functions in a more biologically relevant setting would uncover mechanisms at work that could not be determined with the use of normal laboratory media by developing an *in-vitro* model to determine growth characteristics and the transcriptional profile of *C. difficile* 630 and the mutant strains *Δerm*, *dnaK* and *spo0A*. We have contributed significantly within this thesis to the current understanding of *C. difficile* 630 and many of the processes involved in pathogenesis by providing a deeper understanding of how the transcriptome of *C. difficile* 630 alters in response to faecal water highlighting that important molecular processes are not being picked up in studies that employ basic laboratory growth media. The impact of the faecal water growth environment on *C. difficile* is clear with significant expressional changes observed in sporulation, motility, virulence factors and an important role for regulatory sRNAs in these processes. We have therefore provided a more realistic indication of how this organism may respond in the presence of human gut constituents whilst underscoring the regulatory effect FW has on gene expression in *C. difficile.*

## DECLARATION

I hereby declare that with effect from the date on which the thesis is deposited in Research Student Administration of Ulster University, I permit

1. The Librarian of the University to allow the thesis to be copied in whole or in part without reference to me on the understanding that such authority applies to the provision of single copies made for study purposes or for inclusion within the stock of another library.

2. The thesis to be made available through the Ulster Institutional Repository and/or EThOS under the terms of the Ulster eTheses Deposit Agreement which I have signed.

IT IS A CONDITION OF USE OF THIS THESIS THAT ANYONE WHO CONSULTS IT MUST RECOGNISE THAT THE COPYRIGHT RESTS WITH THE AUTHOR AND THAT NO QUOTATION FROM THE THESIS AND NO INFORMATION DERIVED FROM IT MAY BE PUBLISHED UNLESS THE SOURCE IS PROPERLY ACKNOWLEDGED.

## ABREVIATIONS

| | |
|---|---|
| ANOVA | Analysis of variance |
| BHI | Brain Heart Infusion |
| BHIS | Brain Heart Infusion Supplemented |
| BLAST | Basic Local Alignment Search Tool |
| bp | Base pairs |
| CD | *Clostridioides difficile* |
| CDI | *Clostridioides difficile* infection |
| cDNA | Complementary DNA |
| CDS | Coding sequence |
| Cq | Quantification cycle |
| DNA | Deoxyribonucleic acid |
| dNTP | Deoxyribonucleotide triphosphate |
| dsDNA | Double-stranded DNA |
| DTT | Dithiothreitol |
| EBS | Exon-binding site |
| ELISA | Enzyme-linked immunosorbent assay |
| ErmRAM | Erythromycin retrotransposition-activated selectable marker |
| EtBr | Ethidium bromide |

| | |
|---|---|
| FW | Faecal water |
| g | Gram |
| g/L | Grams per litre |
| GDH | Glutamate dehydrogenase |
| h | Hour |
| HSP | Heat shock protein |
| IBS | Intron-binding site |
| IEP | Intron-encoded protein |
| ƙ | Growth rate constant |
| kDa | Kilodalton |
| mg | Milligram |
| mg/L | Milligram per litre |
| min | Minute |
| ml | Millilitre |
| MLST | Multi-locus sequence typing |
| mm | Millimetre |
| mM | Millimolar |
| mRNA | Messenger ribonucleic acid |
| NCBI | National Centre for Biotechnology Information |

| | |
|---|---|
| ng | Nanogram |
| nL | Nanolitre |
| nm | Nanometre |
| OD | Optical density |
| ORF | Open reading frame |
| PaLoc | Pathogenicity locus |
| PBS | Phosphate buffered saline |
| PCR | Polymerase chain reaction |
| PFGE | Pulse-field gel electrophoresis |
| PMC | Pseudomembranous colitis |
| Pmol | Picomole |
| RAM | Retrotransposition activated marker |
| RIN | RNA integrity number |
| RNA | Ribonucleic acid |
| RNAP | RNA polymerase |
| rpm | Rotations per minute |
| rRNA | Ribosomal ribonucleic acid |
| RT-PCR | Reverse transcription PCR |
| RT-qPCR | Reverse transcription quantitative PCR |

| | |
|---|---|
| s | Second |
| ssDNA | Single-stranded DNA |
| Taq | *Thermus aquaticus* |
| TBE | Tris-Borate-EDTA |
| $T_d$ | Doubling time |
| TF | Trigger factor |
| tRNA | Transfer RNA |
| U | Unit |
| U.K. | United Kingdom |
| U.S.A. | United States of America |
| UV | Ultraviolet |
| V | Voltage |
| v | Volume |
| w | Weight |
| µg | Microgram |
| µl | Microlitre |
| µm | Micrometre |

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF PUBLICATIONS AND PRESENTATIONS

## PUBLICATIONS:

Ternan, N.G., Moore, N.D., Smyth, D., McDougall, G.J., Allwood, J.W., Verrall, S., Gill, C.I.R., Dooley, J.S.G. and McMullan, G. Increased sporulation underpins adaptation of *Clostridium difficile* strain 630 to a biologically–relevant faecal environment, with implications for pathogenicity. *Scientific Reports* (2018) 8: 16691 doi:10.1038/s41598-018-35050-x.

## PRESENTATIONS:

Infection and Immunity Translational Research Group Young Scientist Symposium, Queen's University, Belfast, February 2013.

Poster presentation: Detection of small RNAs in *Clostridium difficile* strain 630.

The Society for General Microbiology Irish Division Autumn Meeting: Gut Microbes - Friend, or Foe?, Ulster University, Coleraine, August 2013.

Poster presentation: Expression of small RNAs in *Clostridium difficile* strain 630.

Infection and Immunity Translational Research Group Young Scientist Symposium, Queen's University, Belfast, February 2014.

Poster presentation: Expression of small RNAs in *Clostridium difficile* strain 630.

# CHAPTER 1: INTRODUCTION

## 1.1 Microbial evolution

Microorganisms are widely believed to be amongst the earliest forms of life on Earth (Dodd *et al*., 2017). Some of the earliest forms of fossilised microorganisms have been found in hydrothermal vent precipitates, from the Nuvvuagittuq belt in Quebec, Canada. They are thought to have formed between 3.8 and 4.3 billion years ago - relatively soon after the formation of Earth some 4.5 billion years ago (Dodd *et al*., 2017). The most common theory amongst evolutionary experts is that the last universal common ancestor (LUCA) to all living things was an anaerobic prokaryote that occupied hydrothermal vents, an environment composed primarily of $H_2$, $CO_2$ and iron (Landan and Graur, 2007; Weiss *et al*., 2016).

Evolution has led to the development of organisms far removed from these primitive microorganisms that inhabited the Earth. However, the development of ribosomal RNA (rRNA) sequencing (Woese *et al*., 1990) has allowed the generation of rRNA phylogenetic trees that help to illustrate the relationships that link all life forms to one another (Figure 1.1) (Woese and Fox, 1977; Woese 1987; Woese *et al*., 1990). In many of the bacterial phylogenetic trees that have been developed, members of the firmicutes are placed close to the root, (and therefore closest to the LUCA core of the tree) with the majority of firmicutes being characterised by their low guanine-cytosine (GC) nucleic acid composition, Gram-positive cell wall and their ability to form endospores (Galperin, 2013; Martin and Sousa, 2016). The firmicutes are generally separated into the anaerobic Clostridia and the obligate or facultative aerobes, including the Bacilli (Galperin, 2013), both of which give rise to a number of important human pathogens including *Bacillus anthracis, Bacillus cereus, Clostridium botulinum* and

*Clostridioides difficile* (previously *Clostridium difficile* (Lawson *et al*., 2016)). *C. difficile* is thought to have emerged approximately 1-2 million years ago but has only been thought of as a human pathogen for just over 40 years (Bartlett *et al*., 1977; He *et al*., 2010; Ramos-Silva *et al*., 2019).

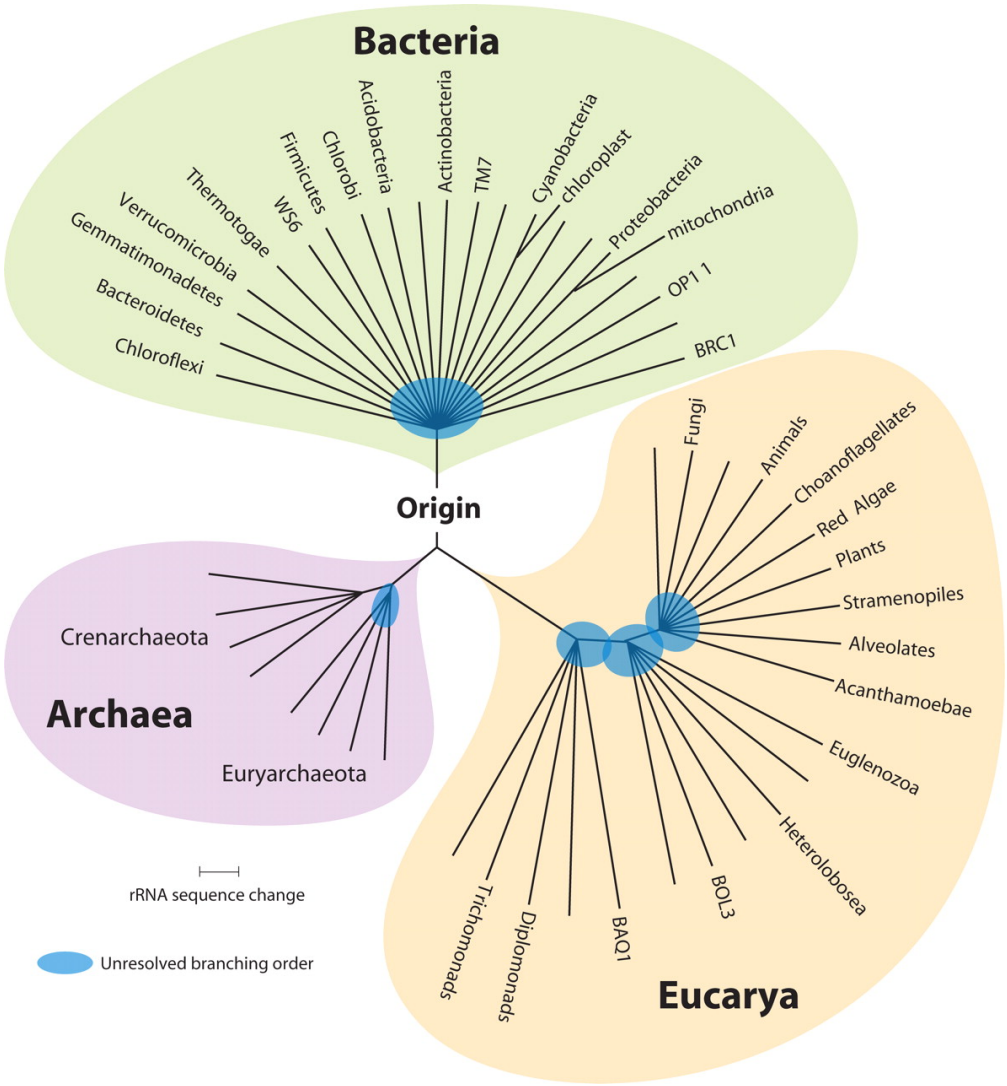



**Figure 1.1:** Phylogenetic tree of life based on genetic sequencing of ribosomal RNA. Origin represents the last universal common ancestor (LUCA). Taken from Pace, 2009.

## 1.2 The Genus *Clostridium*

The bacterial species within the genus *Clostridium* are endospore-producing, low GC-containing, Gram-positive bacteria. They are obligately anaerobic with rod shaped vegetative cells (Sebaihia *et al*., 2006; Galperin, 2013). Members of this species are commonly found in soil, sewage and the intestinal tracts of many animals including humans; and members of this genus have been found to be responsible for causing severe (often life-threatening) disease which can be facilitated in some cases by toxins such as those produced by the organisms *Clostridium tetani*, *Clostridium botulinum* and *Clostridioides difficile.* Illness caused by these organisms can range from food poisoning, infectious diarrhoea and enterocolitis to gas gangrene, tetanus and bacteraemia (Galperin, 2013; Nagahama *et al*., 2019). *Clostridium* species are a vastly phylogenetically varied group of organisms comprising of almost 200 described species, more than 100 of which are spore-forming (Elsayed & Zhang, 2004). Their capacity to form endospores that are resistant to heat, pH and various other chemical perturbations allows them to persist in harsh and often variable environments such as those found in the intestinal tracts of animals.

Many species of *Clostridium* have now had their genomes sequenced and annotated including the pathogenic species *C. perfringens* (Shimizu *et al*., 2002), *C. tetani* (Bruggemann *et al*., 2003), *C. difficile* and *C. botulinum* (Sebaihia *et al*., 2006, 2007). Research has now shown that *C. difficile* is distinctly different to other members of the *Clostridium* genus and therefore it has been placed in a separate genus which has been named "*Clostridioides*" (Lawson *et al*., 2016). 16S rRNA gene sequencing has determined that *C. difficile* is located in the *Peptostreptococcaceae* cluster XI family

of the Clostridia (Yutin and Galperin, 2013). It was initially proposed to change the name of *C. difficile* to *Peptoclostridium difficile* but that has been disregarded in favour of *Clostridioides difficile* as this will allow the commonly used abbreviations *C. difficile* and *C. diff* to continue to be used both medically and commercially (Lawson *et al*., 2016).

Sebaihia *et al*. (2006) were the first to complete the sequence and annotation of the *C. difficile* 630 genome. This particularly virulent strain was first isolated in Zurich, Switzerland, in 1982 from a patient with pseudomembranous colitis (PMC). It was found that more than 1/10th of the genome comprised of mobile genetic elements such as mobile introns, transposons and bacteriophage sequences (Sebaihia *et al*., 2006; Mullany *et al*., 2015). Determining the genomic configuration of these organisms allows further elucidation of the underlying mechanisms leading to virulence and the defence mechanisms employed by the organisms. In addition to this, genomic information offers a greater insight into the usefulness of several members of this genus in the biotechnology industry such as *C. beijerinckii* in the production of acetone and butanol from the metabolism of carbohydrates (Ezeji *et al.*, 2003; Hemme *et al*., 2010; Kalia *et al*., 2011).

### 1.3 *Clostridioides difficile*

*Clostridioides difficile* is a particularly problematic human pathogen - a spore-forming, rod-shaped bacillus which Gram-stains positive and displays obligately anaerobic growth. Found naturally in soil and the intestines of many animals including

humans, it was first discovered in 1935 by Hall and O'Toole when it was isolated from the stool of healthy infants and was originally named to reflect the difficulty encountered when trying to grow it in the laboratory (Hall and O'Toole, 1935). Optimum growth is achieved on blood agar at 37 °C in an anaerobic environment usually made up of approximately 80% $N_2$, 10% $CO_2$ and 10% $H_2$. *C. difficile* grows in cream to yellow ground-glass colonies on agar and are usually 2 mm or more in diameter after approximately 24 h of incubation. They also produce a distinctive horse stable odour (Burnham and Carroll, 2013). Two types of enterotoxins, A and B are produced; these toxins are the main cause of the bacterium's pathogenicity and work to disrupt the host cell's actin cytoskeleton (Di Bella *et al.,* 2016). It wasn't until 1977 that it was discovered to be the cause of diarrhoea and colitis linked to use of the antibiotic Clindamycin (Bartlett *et al*., 1977). Spores are produced when the actively growing organism is placed in an environment it is unable to tolerate, such as extreme pH, temperature, salinity etc. This dormant form allows the organism to survive until favourable conditions are once again encountered, upon which spore germination occurs and active growth can continue (Kochan *et al*., 2018). Toxin and spore production by *C. difficile* will be discussed in greater detail in subsections 1.7 and 1.9, respectively.

### 1.4 *Clostridioides difficile* as a human pathogen

*Clostridioides difficile* infection (CDI) causes infectious bacterial diarrhoea which can lead to life-threatening complications such as pseudomembranous colitis (PMC), a condition characterised by inflammation of the colon and resulting in severe ulceration of the colonic mucosa (Kelly and LaMont, 1998; Bartlett, 2009). Long-term antibiotic

therapy can alter the indigenous microflora of the gut, thereby allowing spores of the organism to germinate; the vegetative cells then grow and produce toxins (see subsection 1.7. Toxin production and role in pathogenesis), which lead to illness (He *et al*., 2010). See Figure 1.2 for an overview of how *C. difficile* leads to disease within the gut. Symptoms resulting from CDI include abdominal pain, cramping, leucocytosis, and a fever of up to 41 °C (Bartlett, 2008). Approximately 80% of CDI cases occur in people over the age of 65, due to the reduction in immune function associated with ageing and the increasing likelihood of co-morbidities and polypharmacy in these individuals (Bartlett, 2009). The anoxic environment of the inside of the colon is the ideal setting for anaerobic organisms such as *C. difficile* to survive. Thereby, colonisation in other areas of the body including the small intestine is uncommon (Abt *et al*., 2016).

CDI often starts off as simple water-dense diarrhoea, which can then progress towards more serious issues such as colonic inflammation and the formation of lesions within the colonic mucosa. At an advanced stage of the infection, pseudomembranous colitis and toxic megacolon, which is characterised by severe distension of the large bowel, can occur; both of which are indicative of CDI. In a fifth of cases of advanced disease patients experience very little diarrhoea and instead present with abdominal distention and ileus, in these cases misdiagnosis as irritable bowel syndrome (IBS) or inflammatory bowel disease (IBD) is common (Ramsey *et al*., 2002; Jackson *et al*., 2016). In contrast, systemic inflammatory response syndrome (SIRS) can develop in many instances of CDI, leading to severe hypotension, acute renal failure, and respiratory distress; in these cases the prognosis is often poor (Rao *et al.* 2014). The

development of fulminant colitis often requires total colectomy and intensive antibiotic therapy. However, despite these measures mortality rates remain at approximately 70% in these patients (Sartelli *et al,* 2015; Ramsay *et al*., 2018). It can be difficult to predict when or if CDI will lead to fulminant CDI requiring surgery. Progression can be within hours of symptom onset to weeks from initial presentation. Many factors are thought to be involved such as patient age, immune status, co-morbidities, current medication and the virulence of the strain of *C. difficile* involved (Heinlen and Ballard, 2010; Asempa and Nicolau, 2017).

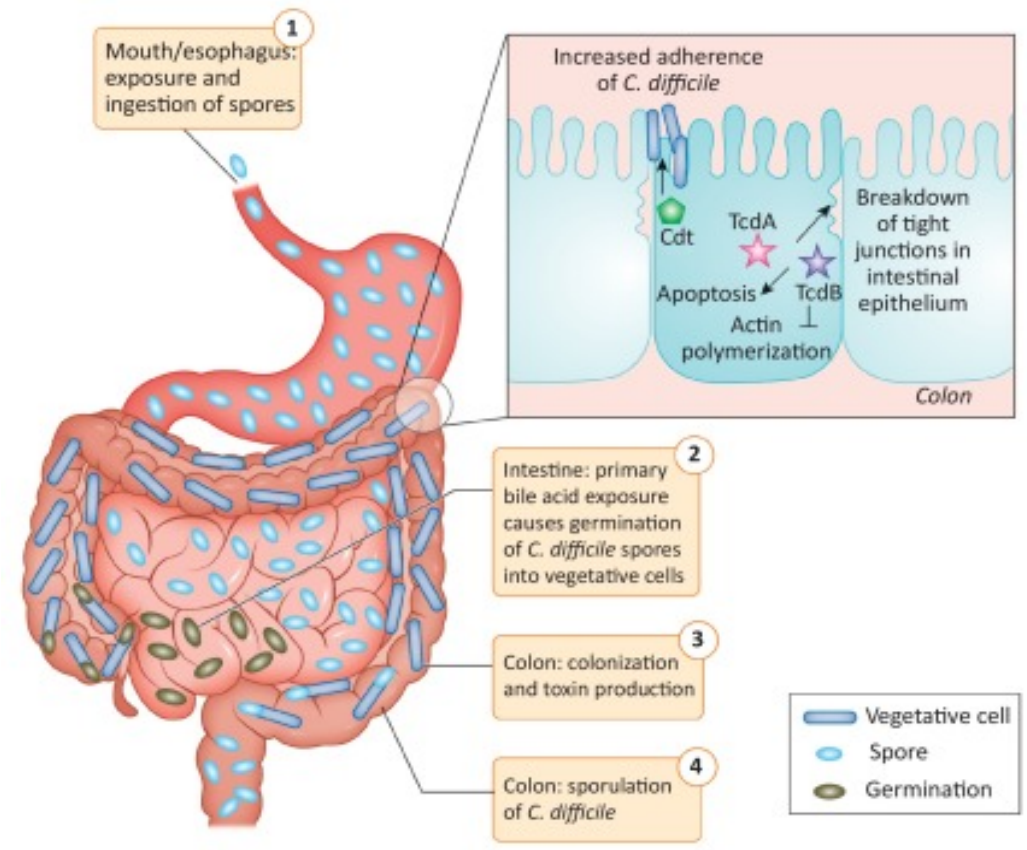



**Figure 1.2:** Development of *Clostridioides difficile* infection (CDI) within the human intestine (Sandhu and McBride, 2018).

## 1.5 Epidemiology of *Clostridioides difficile* infection

The collection of epidemiological data allows clinicians and researchers to track, catalogue and compare pathogenic strains. This information is important as it enables the planning for and control of public health concerns such as an outbreak of *C. difficile* infection within a community or healthcare setting. The more that is known about the factors that contribute to disease outbreaks the better equipped we are to reduce the severity of any future events (Martin *et al*., 2016). It has been reported that *C. difficile* is the cause of up to 30% of all incidents of infectious diarrhoea associated with antibiotic consumption, with approximately half a million patients in US hospitals in 2006 being diagnosed with the condition (Rupnik *et al*., 2009; Khan and Elzouki, 2014). An increase in the number of cases of CDI across the United States, Canada and Europe over the last two decades has been associated with the emergence of the so called hypervirulent strain of *C. difficile* BI/NAP1/027, first identified in Quebec, Canada in 2005 (McDonald *et al*, 2005; Tillotson and Tillotson, 2011; Balsells *et al*., 2019). The virulent nature of this strain has been linked to the increased production of the enterotoxins A and B and the additional virulence factor, binary toxin (*C. difficile* transferase (CDT)), in addition to increased resistance to the fluoroquinolones. By 2010 this strain had spread worldwide (McDonald *et al*, 2005; Pépin *et al*., 2005; Clements *et al*., 2010). Increased surveillance within Europe including the UK and Northern Ireland between 2004 and 2008 identified ribotypes 027, 106 and 001 as the most common with deaths from the illness rising by 400% during this period within England and Wales (Tillotson and Tillotson, 2011). An outbreak caused by ribotype 027 occurred at the Stoke Mandeville Hospital in 2003-2004 followed later that year by a second outbreak causing a total of 38 deaths. An investigation was carried out which found that there was a lack of isolation facilities available and that staff were

not following adequate infection control practices such as hand-washing and changing PPE often enough as they did not have the time due to a staff shortage (Commission for Healthcare Audit and Inspection, 2006).

The spread of CDI has also been problematic in long-term healthcare facilities such as care homes for the elderly due to a number of factors including advanced age, immune compromisation and the communal nature of the living environment (Freeman *et al*. 2010). A study by Khanna *et al*. (2012) has found that in comparison to healthcare acquired CDI, community acquired is more likely to occur in the younger population (median age of 50 years compared to 72 years in hospital acquired) and is more prevalent than previously thought. McDonald *et al.* (2007) define community acquired CDI as diarrhoea occurring where the patient has not been admitted to hospital in the preceding 12 weeks or for symptoms to occur within 48 h of admission to hospital. Community acquired CDI is still not as widely recognised as nosocomial CDI (Al Assaad *et al*., 2020) and this has led to a lack of screening for the disease in addition to the increasing use of broad-spectrum antibiotics and gastric reflux medications such as proton-pump inhibitors and $H_2$ receptor blockers in general practice – all of which are known to contribute to the development of the disease. The elevation in gastric pH caused by gastric reflux medications reduces the bactericidal effect of the gastric acid, increasing the chance of *C. difficile* colonisation (Kyne *et al* in 1998; Kwok *et al*., 2012; Janarthanan *et al*., 2012; Freedberg and Abrams, 2013).

Rates of deaths related to *C. difficile* infection have been declining within Northern Ireland in recent years. Most up-to-date statistics from the Northern Ireland Statistics

and Research Agency (NISRA) show that in 2017 there was a total of 52 deaths where *C. difficile* was mentioned on the death certificate compared to 191 in 2008 (Figure 1.3). This decline can most likely be to attributed to increased awareness amongst both healthcare staff and the public in addition to the implementation of more robust infection control mechanisms (see subsection 1.11. Treatment and prevention for further information on infection control mechanisms).

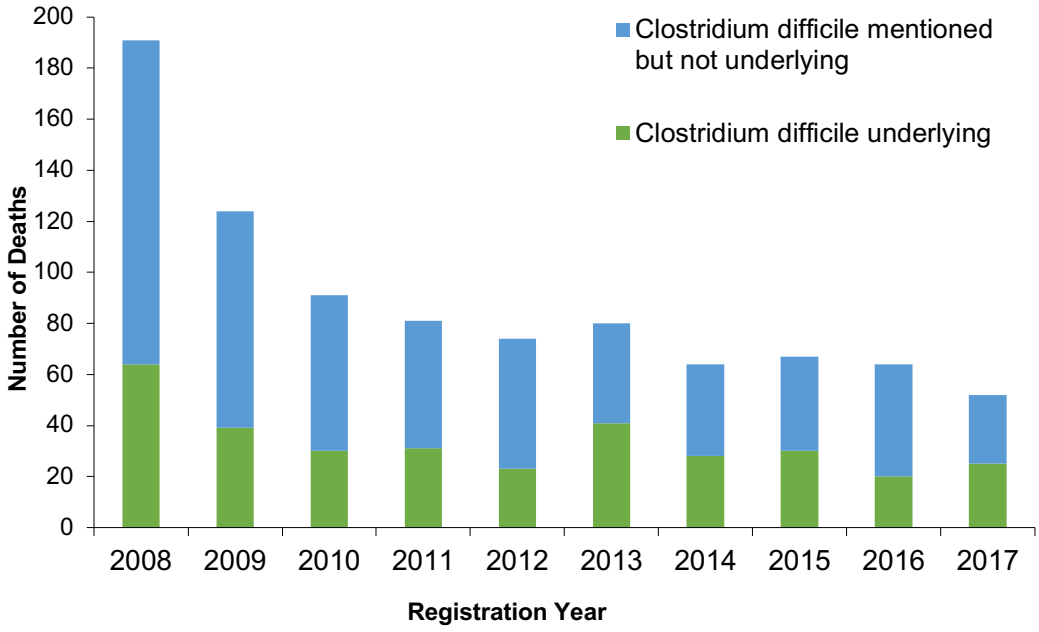



**Figure 1.3:** Mortality rates for deaths involving *Clostridioides difficile* registered in Northern Ireland between 2008 and 2017 (Northern Ireland Statistics and Research Agency (NISRA), December 2018).

## 1.6 Typing methods

The identification of Clostridia by established procedures such as biochemical testing and conventional culture is still costly and time-consuming (Humphries and Linscott, 2015). Moreover, many commercially available methods for the identification of anaerobic microorganisms phenotypically, have proven unsuccessful in correctly recognising *Clostridium* species (Planche *et al*., 2008; Bagdasarian *et al*., 2015; Humphries and Linscott, 2015). These drawbacks have led to the development of newer, more efficient molecular techniques to aid in the identification of many important species within the genus (Song, 2005; Lindstrom and Korkeala, 2006). A number of typing techniques have been developed to allow identification and differentiation between isolates, thereby allowing the epidemiology of *C. difficile* to be investigated and monitored (Curry, 2010; Griffiths *et al*., 2010). Conventional typing methods that were formerly based on phenotypic properties of the various strains have since been superseded by more advanced molecular methods such as PCR ribotyping.

### 1.6.1 PCR Ribotyping

PCR ribotyping is based on identification of disparities in the nucleotide sequence of the 16S and 23S ribosomal RNA (rRNA) intergenic spacer regions of *C. difficile* strains and was first described by Gurtler *et al*. (1991). The amplified nucleic acid is cut using restriction enzymes producing bands of varying patterns. These bands are compared to reference libraries in order to identify the "ribotype". Compared to many other bacterial species, *C. difficile* has significant intraspecific variety in the intergenic

spacer regions making it particularly suitable to PCR ribotyping (Sadeghifard *et al*., 2006).

### 1.6.2 Restriction endonuclease analysis (REA)

REA uses whole-genome DNA, which has been digested by the restriction enzyme *Hind*III. Agarose gel electrophoresis is then used to separate the resultant restriction fragments. The advantages of this technique are that it is stable and highly discriminatory. However, the process can be technically arduous and the data is not easily compared between different laboratories (Huber *et al*., 2013).

### 1.6.3 Surface layer protein A-encoding gene (s*lpA*) typing

It has been found that the surface layer protein (SlpA) varies amongst *C. difficile* strains, therefore sequencing of the variable region of the *slpA* gene has been utilised as a method for typing *C. difficile* (Mori and Takahashi, 2018). This method is not widely used as it has been shown that isolates with the same ribotype or MLST type have different *slpA* types (Dingle *et al*., 2013).

### 1.6.4 Arbitrarily primed PCR AP-PCR

An alternative molecular technique known as arbitrarily primed PCR (AP-PCR) identifies polymorphisms within the *C. difficile* genome without the requirement of advanced knowledge of the target nucleotide sequence (Brazier, 2001). Although rapid and simple, reproducibility for this technique can be poor (Bidet *et al*., 2000).

### 1.6.5 Toxinotyping

*C. difficile* toxin production can also be used to distinguish between the strains, this technique is known as toxinotyping and utilises restriction fragment length polymorphism (RFLP)-PCR, by identifying sequence variations in the pathogenicity locus (PaLoc) (a region of the genome containing toxins A and B genes), which are compared to a reference strain. Each strain is categorised into one of 27 toxinotypes, with *C. difficile* 630 classed as toxinotype 0. Toxinotyping is useful for identifying the toxin status of the strain (i.e. which combination of A, B and/or binary toxin the strain possesses) although it does not have the same discriminatory power as other methods such as PCR ribotyping (Rupnik *et al*., 2001; Huber *et al*., 2013; Singh *et al*., 2017).

### 1.6.6 Multilocus sequence typing (MLST)

Multilocus sequence typing (MLST), enables the rapid genotyping of *C. difficile* isolates through the amplification of reference genes such as those that code for triosephosphate isomerase (*tpi*), recombinase A (*recA*) and adenylate kinase (*adk*) (Griffiths *et al*., 2010). MLST is useful for population studies but is relatively expensive compared to PCR ribotyping (Kuijper *et al*., 2009).

### 1.6.7 Multilocus variable-number tandem-repeat analysis (MLVA)

Capillary electrophoresis is utilised in MLVA to resolve fragments by size, and determine the number of repeat sequences that are present. MLVA has more discriminatory ability than PCR ribotyping and is able to establish phylogenetic

relationships between strains. This technique can generate digital data that can be shared between labs, however it can be expensive to carryout (Hardy *et al*., 2012).

### 1.6.8 Pulsed-field gel electrophoresis (PFGE)

PFGE separates large fragments of whole-genome DNA which have been cut using restriction enzymes such as *Sma*I or *Sac*II. Some of the first forms of this method were unreliable due to DNA degradation and difficulties with lysing spores. However, modifications to the electrophoresis method have led to the ability of this technique to achieve 100% typeability. One major disadvantage of this method is speed, with complete analysis taking several days to obtain (Huber *et al*., 2013).

### 1.6.9 Whole genome sequencing

Whole genome sequencing (WGS) allows detailed characterisation of pathogens in all kinds of samples using next generation sequencing (NGS) technology (Sabat *et al*., 2017). Advantages include exact strain determination allowing tailored patient management and accurate pathogen surveillance data, which can be quickly shared between laboratories for monitoring purposes (Rossen *et al*., 2018). WGS is being increasingly used in clinical microbiology laboratories however a number of issues prevent it from replacing routine microbiology at the present time including turnaround time, cost and complications of downstream data analysis, which must be performed by highly trained personnel (Rossen *et al*., 2018). Although many methods have now been developed, PCR ribotyping remains the most straightforward,

discriminatory and affordable method available (Brazier, 2001; Griffiths *et al*., 2010; Huber *et al*., 2013).

### 1.7 Toxin production and role in pathogenesis

Whilst numerous virulence factors are known to be involved in *C. difficile* pathogenesis and colonisation, the development of CDI is strongly associated with the presence of two toxins - Toxin A (TcdA) and toxin B (TcdB) (Awad *et al.,* 2014; Chandrasekaran and Lacy, 2017). Toxin A is a 308 kDa enterotoxin and toxin B is a 270 kDa cytotoxin, both exert tissue damaging ability, attacking the colonic mucosa through binding to membrane receptors and entering the mucosal cells by receptor-mediated endocytosis (Genth *et al*., 2008). A 19.6 kb chromosomal region known as the pathogenicity locus (PaLoc) (Figure 1.4) is where the Toxin A (*tcdA*) and B (t*cdB*) genes are located within the *C. difficile* genome. The PaLoc consists of *tcdA* and *tcdB*, as well as three other genes involved in the regulation of toxin gene expression – *tcdR*, *tcdE*, and *tcdC*. *TcdE* encodes a holin-like protein, a type of bacteriophage protein involved in the release of progeny phages from the host organism. Consequently, TcdE is also thought to play a role in the release of TcdA and TcdB through the cell wall (Tan *et al*., 2001; Govind and Dupuy, 2012). *TcdR* encodes an extra-cytoplasmic sigma factor which works to positively regulate itself in addition to *tcdA* and *tcdB* (Mani and Dupuy 2001; Mani *et al.,* 2002). Sigma factors help to facilitate the binding of RNA polymerase to promoter regions within the genome to enable the initiation of transcription of specific genes in response to the cellular environment. They join together with RNA polymerase to form the RNA polymerase holoenzyme (Paget, 2015). Conversely, *tcdC* has been found to negatively regulate toxin gene expression,

behaving mainly as an anti-sigma factor by inhibiting transcriptional activity (Dawson, *et al*., 2009; Carter *et al.,* 2011).

Researchers have uncovered strains of *C. difficile* where regions of the PaLoc are missing, these are referred to as A-B+ or A-B- strains (Dawson, *et al*., 2009). One such study by Stabler *et al*. (2006) looked at 75 strains of *C. difficile* and identified four separate clusters – a hypervirulent (HY), toxin-variable (A-B+), and two human/animal clades (HA1 and HA2). Differences separating the clades included variations in toxin, antibiotic resistance, adhesion, motility and metabolism gene sequences (Stabler *et al*., 2006). It has also been noted that non-toxin producing *C. difficile* strains can acquire the PaLoc from toxigenic strains by horizontal gene transfer, thereby transforming non-toxigenic strains into toxin expressers (Brouwer *et al.,* 2013). This gene transfer is mediated through the integration of active conjugative transposons (CTns) into the genome. CTns are genetic elements that are present in toxin producing strains such as *C. difficile* 630 and have been shown to be capable of inserting themselves into other strains such as the non-toxin producing *C. difficile* CD37 (Brouwer *et al.,* 2013). The conjugative transposon CTn*2* was shown to be able to transfer the PaLoc from *C. difficile* 630 to strain CD37 and more importantly this strain was shown to be able to express the acquired toxins (Brouwer *et al.,* 2013). As there have been proposals to use non-toxigenic strains as a means of treatment of CDI this finding is of great significance (Villano *et al*., 2012).

A binary toxin known as CDT has been found in a number of strains, initially in the CD196 strain (non-epidemic PCR-ribotype 027 strain) isolated from a patient in Paris in 1985 (Popoff *et al.*, 1988). CDT is made up of an enzymatic component and a binding component (CdtA and CdtB, respectively) and works mainly to cause disruption of the host epithelial cell cytoskeleton, with CdtA modifying actin (by enabling ADP-ribosylation of monomeric actin causing depolymerisation of actin filaments) and CdtB helping CdtA cross into the cytosol. CdtB essentially acts as a transport element by facilitating the translocation of CdtA into the target cells (Davies *et al*., 2011). It is also believed that the presence of the binary toxin and an associated regulatory protein (CdtR) may heighten the toxicity of TcdA and TcdB by positively regulating the expression of *tcdA* and *tcdB*, thereby increasing virulence of the organism and leading to more severe disease (Denève, *et al*., 2009).

The virulence of the epidemic *C. difficile* ribotype 027 strain is thought to be due to a frameshift mutation present in the *tcdC* coding sequence at nucleotide position 117, with this polymorphism leading to the production of a truncated protein (Wolf *et al*., 2009). Consequently, TcdC is unable to carry out its role in negatively regulating toxin production, causing increased expression of toxin A and B by approximately 16-fold and 23-fold, respectively, when compared to C. *difficile* isolates that lack the frameshift mutation (Warny *et al*., 2005). Although toxins are the main virulence factors, others do exist such as adhesins including GroEL, cell wall proteins and flagellar proteins (Hennequin *et al*., 2003; Sebaihia *et al*., 2006; Janoir *et al*., 2007; Fagan *et al*., 2009).

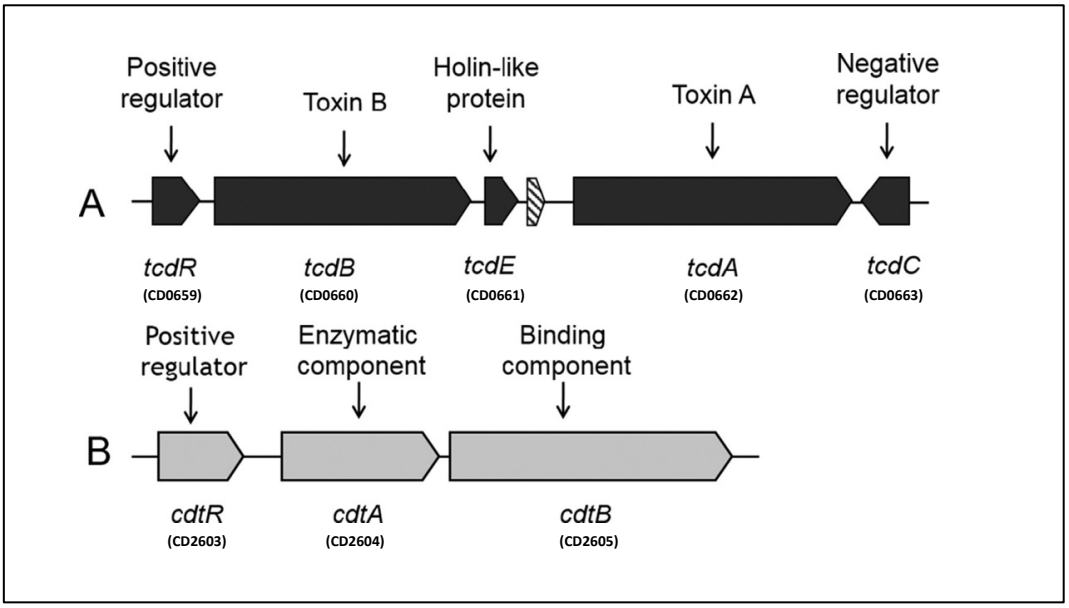



**Figure 1.4:** Two main *C. difficile* toxin loci. (A) The PaLoc encodes toxin A (*tcdA*) and toxin B (*tcdB*), as well as three accessory proteins (*tcdR*, *tcdE* and *tcdC*). Dashed arrow represents a partial pseudogene found downstream of *tcdE*. (B) CdtLoc encodes two binary toxin genes (*cdtA* and *ctdB*) and one accessory gene (*ctdR*) (Elliot *et al*., 2017).

### 1.8 Non-toxin virulence factors

Bacterial organisms such as *C. difficile* are able to survive during exposure to stressful conditions in a number of ways. One such mechanism is the heat shock response in which molecular chaperones, called heat shock proteins (HSPs), are responsible for conferring protection to the organism by refolding damaged proteins. HSPs respond not only to heat but also to many other stressful situations such as nutrient deprivation or antimicrobial therapy. Numerous molecular chaperones have been identified including HSP60, HSP70, HSP90 and HSP100 amongst many others. The molecular chaperone GroEL, a member of the HSP60 family, has been shown to be involved in the binding of cells to the epithelial lining of the intestine and is found at increased

levels at the surface of *C. difficile* cells heat-shocked at 48 °C (Hennequin *et al*., 2001a; 2001b; 2003; Jain *et al*., 2011; Jain *et al*., 2017). Many environmental conditions can lead to an increase in gene expression and protein abundance of this particular molecular chaperone such as increased temperature and osmolarity, iron deficiency, and antibiotics (Hennequin *et al*., 2001a; Emerson *et al*., 2008). A member of the HSP70 family of proteins, DnaK, has been shown to assist in the folding of non-native polypeptides (i.e. polypeptides that are not in their correct 3D form and therefore lack function (Liberek *et al*., 2008). Jain *et al*. (2017) found that biofilm formation is increased in a *dnaK* ClosTron disruption mutant when compared to the wildtype, suggesting that this increase results from the response of the organism to stress and is used as a protection mechanism. Furthermore, the *dnaK* mutant did not form observable flagella thereby rendering the organism immotile (Jain, 2010; Jain *et al*., 2017; Alotaibi, 2017).

These observations, in addition to the *dnaK* mutant's sensitivity to elevated temperatures and impaired growth rates, indicate that the organism is, physiologically, in a state of heat stress due to lack of DnaK protein function (Jain *et al*., 2017). These significant observations in the *dnaK* disruption mutant highlight the importance of molecular chaperones in the functioning of virulent and pathogenic strains of *C. difficile*.

## 1.9 Sporulation

*C. difficile* is a spore-forming organism meaning that it possesses the ability to produce a highly resistant, infectious form of itself when under stressful conditions. Spores allow spore-producing organisms to survive when placed in an environment not normally conducive to sustained growth until they again find themselves in more favourable conditions when the vegetative form can proliferate. Sporulation is one of the main reasons *C. difficile* is so highly contagious, as spores are easily transmitted from the faeces of both humans and animals (Brown and Wilson, 2018). Sporulation takes place in the gut, subsequently spores are shed in the faeces – transmission from person to person then ensues *via* the faecal-oral route (Freeman *et al*., 2010; Zhu *et al*., 2018).

Sporulation in *C. difficile* involves a complicated sequence of stages ultimately leading to the development of metabolically inactive spores from vegetative cells (Figure 1.5) (Zhu *et al*., 2018). In *C. difficile*, the resilient nature of the spores can be accredited to their structure, consisting of an inner core containing the complete genome, which is surrounded by the cortex (made-up of modified cross-linked peptidoglycans), this is further encircled by the spore-coat composed of a thick coat of peptidoglycan and a final outer layer known as the exosporium – a loose-fitting structure composed of a basal layer and an outer layer of hair-like projections (Paredes-Sabja *et al*., 2014). The exosporium is not present in all spore-forming bacteria but is found in many clostridial species including the pathogens *C. botulinum* and *C. difficile* amongst others (Paredes-Sabja *et al*., 2014). Sporulation in *C. difficile* typically follows 4 main stages. Stage 0 represents normal vegetative cell replication by binary fission. During Stage I

asymmetric septation occurs creating a smaller compartment (SC) and a larger mother cell (MC). Stage II is where the SC becomes the forespore and is engulfed by the MC in a phagocytic manner. In Stage III construction of the spore cortex and coat begins followed by Stage IV when the MC breaks down to discharge the mature spore into the immediate environment (Fimlaid and Shen, 2015; Zhu *et al*., 2018).

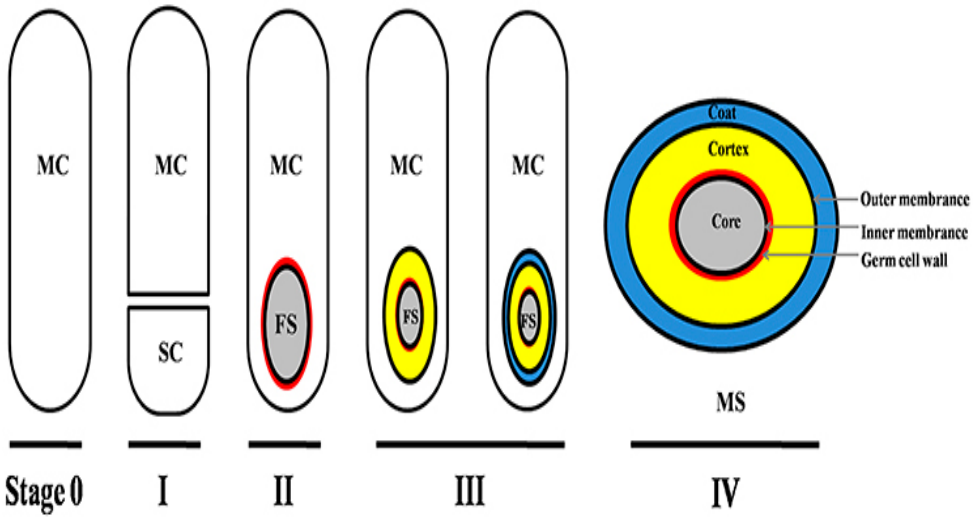



**Figure 1.5:** Key morphogenic stages of sporulation and spore structure in *C. difficile*. MC, mother cell compartment; SC: smaller compartment; FS: forespore compartment; MS: mature spore (Zhu *et al*., 2018).

At the transcriptional level, *C. difficile* sporulation is initiated by five orphan histidine kinases (a histidine kinase gene not directly associated with a response regulator gene), CD1352, CD1492, CD1579, CD1949 and CD2492, when environmental signals indicate that growth is no longer feasible. This leads to phosphorylation of the master transcriptional regulator Spo0A; which controls RNA polymerase sigma factors σF,

σE, σG, and σK (Underwood *et al*., 2009; Steiner *et al*., 2011; Fimlaid and Shen, 2015). CD1579 is the only kinase known to phosphorylate Spo0A directly (Underwood *et al*., 2009). The σ factors σF, σE, σG, and σK have been found to control transcriptional regulation in a compartment-specific manner. Factors σE and σK are specific to the mother cell (MC), whereas σF and σG are specific to the developing forespore (FS) (Pereira *et al*., 2013; Al-Hinai *et al*., 2014; Fimlaid and Shen, 2015; Zhu *et al*., 2018). In contrast to *B. subtilis* the activity of σE in *C. difficile* is not fully dependent on σF. In addition to this, σG does not require σE for activation, and σK does not require σG for activation. Moreover, the post-translational activation of σG was found to be dependent on σF (Fimlaid *et al*., 2013). These four sigma factors interact with a number of sporulation-associated genes in a controlled systematic fashion usually in response to either internal or external (e.g. bile acids, nutrient deprivation etc.) stimuli; see Figure 1.6 for a detailed overview of the proposed sporulation cascade in *C. difficile* by Pettit *et al*. (2014).

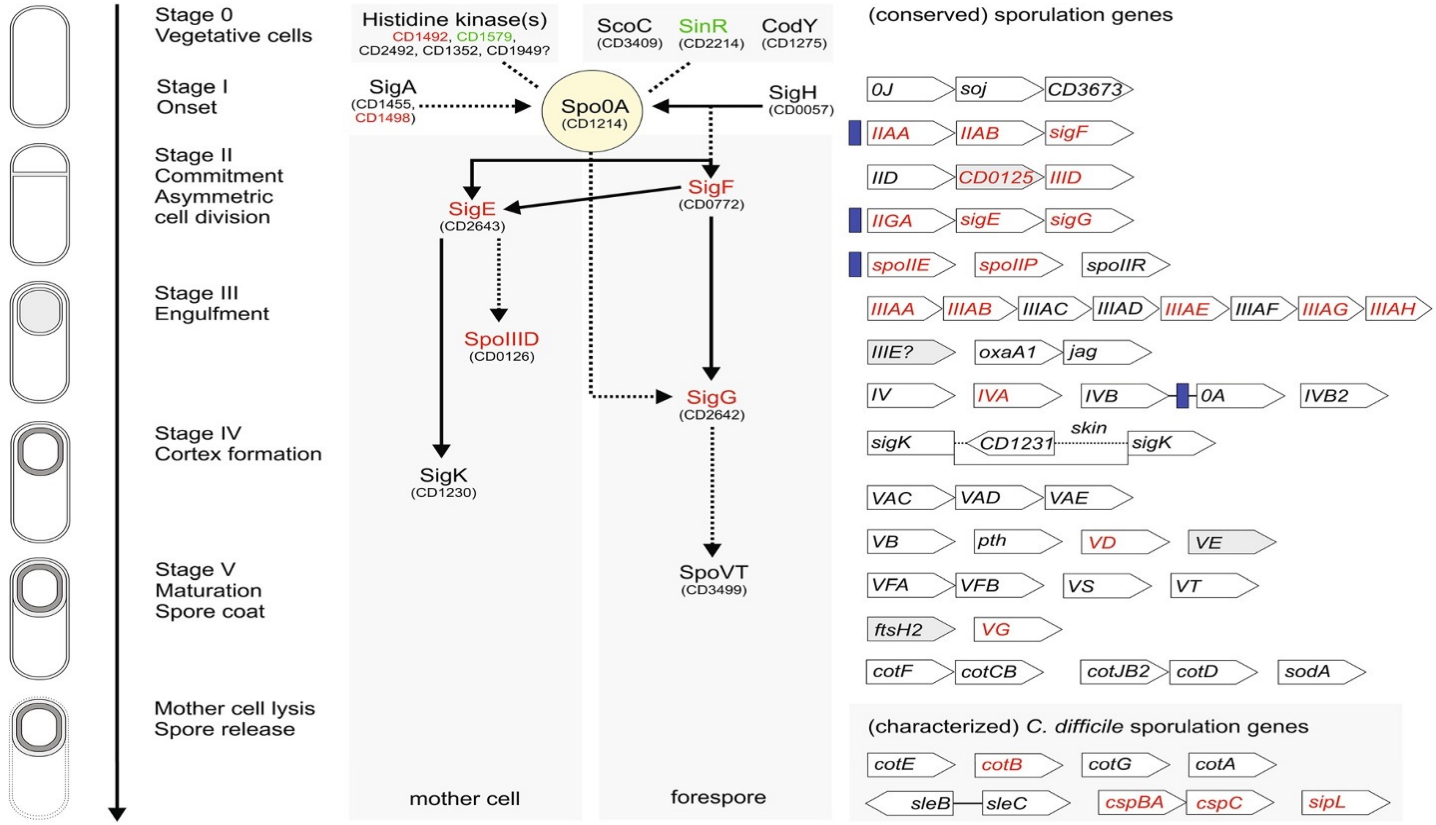



**Figure 1.6:** Proposed sporulation cascade for *Clostridioides difficile* 630 developed by Pettit *et al*. (2014). Solid arrows in the putative regulatory cascade of *C. difficile* are confirmed interactions. Dotted arrows are uncharacterized or proposed interactions. Vertical blue bars are *in vitro* confirmed Spo0A binding sites. Genes in grey represent tentative homologs of *B. subtilis* sporulation genes. Green colours specify genes upregulated and red colours specify genes downregulated in a *spo0A* mutant compared to parental *C. difficile* 630 *Δerm* strain (Pettit *et al*., 2014).

### 1.10 Diagnosis of *C. difficile* Infection (CDI)

Testing for CDI is carried out when a patient presents with typical symptoms associated with CDI such as abdominal pain and diarrhoea and has recently undergone a course of antibiotic therapy (short or long term). In some cases, they may also have a fever of up to 41 °C (Bartlett and Gerding, 2008). The main determinants of a suitable test for the diagnosis of CDI are cost, availability, specificity, sensitivity and turnaround time (Gilligan, 2015). There are five tests that are currently accepted as being clinically useful. They are glutamate dehydrogenase (GDH) detection, enzyme immunoassay (EIA) for detection of toxins A/B, toxigenic culture (TC), cytotoxin neutralization (CTN) and nucleic acid amplification tests (NAATs) (Avila *et al*., 2016). The glutamate dehydrogenase (GDH) test uses antibodies to detect the presence of the GDH enzyme and is a highly sensitive immunoassay method that allows reliable exclusion of *C. difficile* if the result is negative (Fenner *et al*., 2008). If a positive GDH result is produced, further testing is recommended, such as EIA toxin test for the presence of toxins A/B (Crobach *et al*., 2016; Martinez-Melendez *et al*., 2017). EIA tests tend to produce a signal on detection of the toxin, usually this is a colour change and can be measured spectrophotometrically via a microplate reader (Crobach *et al*., 2016). Unfortunately, most commercially available toxin EIAs lack sensitivity, at 60-95% sensitive, which is why nucleic acid amplification tests (NAAT) are more often used to confirm the results of the EIA. NAATs can provide a more rapid diagnosis compared to GDH and EIA. However, false positives can occur due to the sensitivity of the NAATs and their inability to distinguish between active toxins as they are only able to detect the presence of the toxin gene. This can lead to over-diagnosis and potentially unnecessary treatment (Avila *et al*., 2016).

Real-time polymerase chain reaction (RT-qPCR) can be utilised in order to rapidly confirm the presence of specific genes such as the *tcdB* gene which controls production of toxin B or the gene which negatively regulates toxin A and B production, *tcdC*. This method can also be used to detect nucleotide deletions in the *tcdC* gene, known to cause over-production of toxins A and B in the B1/NAP1/027 *C. difficile* strain (discussed in subsection 1.7. Toxin production and role in pathogenesis), providing the added benefit of being able to distinguish between strain types (Crobach *et al*., 2016). Various algorithms of these methods are often employed in different laboratories making the diagnosis of CDI far from straight forward (Ignatius *et al*., 2019). In the United Kingdom the guidelines issued by the Department of Health recommend a combination of two tests should be carried out to confirm the diagnosis of CDI. It is recommended that the initial test is a NAAT or GDH followed by a toxin EIA test (ARHAI, 2012). Although guidelines should be followed as closely as possible for all clinical decision making, the choice of testing algorithm used is ultimately laboratory and patient specific (Avila *et al*., 2016).

### 1.11 Treatment and prevention

First-line treatment for *C. difficile* infection (CDI) should involve the immediate removal of the initiating antimicrobial agent and replacement with antibiotics that are less commonly associated with the development of CDI such as the imidazoles (e.g. metronidazole), glycopeptides (e.g. vancomycin) and macrolides (e.g. fidaxomicin) (Bauer *et al*., 2009; Debast *et al*., 2014). In mild cases, withdrawal of the causative antimicrobial can often lead to cure of CDI without further treatment (NHS, 2018). Vancomycin tends to be used secondary to metronidazole due to growing fears over

the development of vancomycin resistant Enterococci (VRE) and thus, it is reserved for patients who are intolerant to, or fail to respond to metronidazole (Tacconelli and Cataldo, 2008; Faron *et al*., 2016). Metronidazole and vancomycin have been in use as therapeutic agents against CDI for over three decades and are still first-line options. Current therapeutic targets for CDI include DNA, RNA, protein synthesis and cell wall integrity amongst others (see Figure 1.7 for overview). In addition to antimicrobial therapy, treatment plans should include the implementation of strict infection control policy and the correction of fluid loss and electrolyte imbalance (Leffler and Lamont, 2015; Gerdling *et al*., 2016; Napolitano and Edmiston, 2017).

Under stressful conditions such as heat, antibiotics or nutrient limitation, spores are produced that are highly resistant to various environmental conditions including antibiotics and many sterilisation methods such as industrial cleaners, making them difficult to eradicate (Sorg and Sonenshein, 2008; Kochan *et al*., 2018). The antibiotics required to eradicate the vegetative bacteria do not kill the spores and often contribute to recurrence by disrupting the normal gut flora even further. In the event that antibiotic therapy fails or exacerbates the infection, a colectomy may be carried out in severe, complicated or fulminant cases of CDI as a very last resort to reduce the potential for progression of the colitis and development of toxic megacolon (Bauer *et al*., 2009; Surawicz *et al.*, 2013; Gerdling *et al*., 2016).

In instances of recurrent CDI, oral vancomycin or fidaxomicin is recommended. Faecal microbiota transplantation (FMT) has also been found to be an effective

treatment option in recurrent cases, whereby the normal gut microflora is restored with faecal matter from a donor, administered most commonly via nasoduodenal tube. FMT has been shown to be not only safe but also successful in treating recurrent CDI in over 90% of patients (Kassam *et al*., 2013). Widespread use of FMT is held back by the need for regulated stool banks to safely distribute donor stool to hospitals when required. Dedicated centres are essential to ensure safety and traceability of donations in addition to monitoring of outcomes for quality control purposes (Cammarota *et al*., 2019). As rates of antimicrobial resistance continue to rise this particular form of therapy is likely to become more widely utilised in the treatment of intestinal infections.

Strict implementation of measures to prevent the spread of *C. difficile* are imperative in the healthcare environment. The use of gloves and gowns on entry to the room of a patient and stringent hand washing by all hospital staff and visitors as well as sterilisation of surfaces and equipment are all important measures in the prevention of transmission (Gerding *et al*., 2008, Hsu *et al*., 2010). Public Health England strongly recommend immediate isolation of suspected CDI cases, alongside prompt treatment with appropriate antibiotics to prevent further spread and reduce the possibility of an outbreak from occurring (Wilcox, 2013; Peng *et al.*, 2018). A potentially complicating factor is the asymptomatic carriage of *C. difficile* which is thought to affect between 7 - 15% of the population (Galdys *et al*., 2014) and which is transient in nature, making strict infection control practices all the more important (Furuya-Kanamori *et al*., 2015). However, careful antibiotic prescribing practices are by far the most important factor in the control and prevention of CDI (Smits *et al*., 2016). Very few antibiotics

remain that haven't been implicated in the development of CDI, but those most commonly associated are amoxicillin, fluoroquinolones, cephalosporins, ampicillin and clindamycin. Reducing the number and duration of prescriptions of these antibiotics (antibiotic stewardship) remains key to controlling the development and spread of this pathogen (Surawicz *et al*., 2013).

The use of probiotics has been investigated extensively in recent years with widely varying outcomes. Many studies have concluded that there is a benefit to the use of probiotics including various lactobacilli species such as *L. acidophilus*, *L. casei*, *L. rhamnosus* and *L. plantarum* in the prevention of CDI (Gao *et al.,* 2010; Johnston *et al*., 2012; Maziade *et al.,* 2015; Ratsep *et al.,* 2017) while others have determined that probiotics are ineffective as a prophylactic against CDI (Pochapin, 2000; Wullt *et al.,* 2003; Pozzoni *et al*., 2012; Allen *et al*., 2013). For example, the use of *Bacteroides fragilis* as a barrier to *C. difficile* colonisation was demonstrated by Deng *et al*. (2018) who showed in a mouse model that pathogen colonisation was decreased, and gut barrier integrity and function was increased, thereby preventing epithelial stress and destruction. The study of novel targets for the treatment of *C. difficile* has become increasingly important in this era of growing antibiotic-resistance (Figure 1.7).

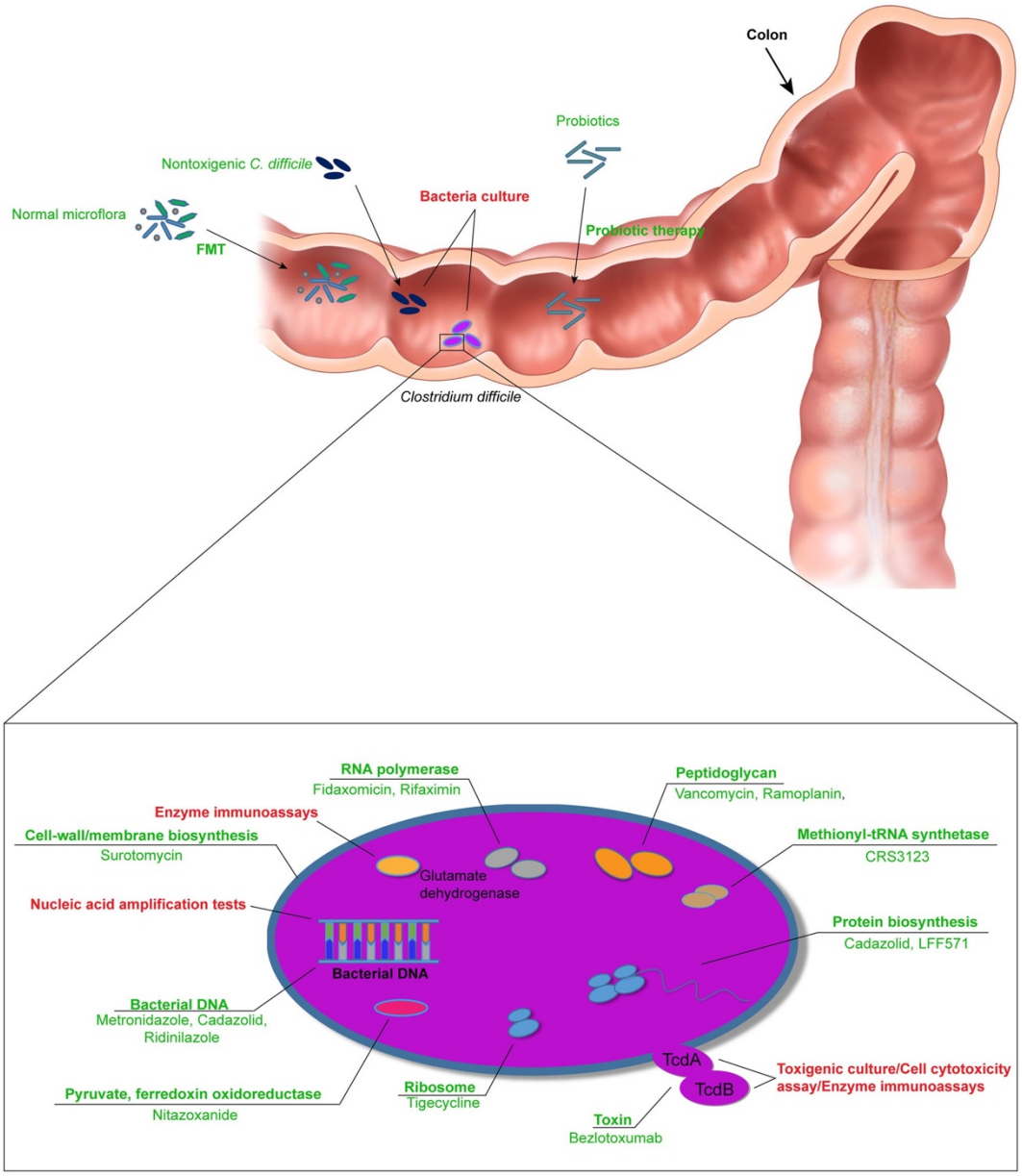



**Figure 1.7**: Bacterial targets for *Clostridioides difficile* infection diagnosis and treatment. Text in red - indicated targets for diagnostics. Text in green - indicated targets for therapeutics (Peng *et al.*, 2018).

### 1.12 Gene editing – ClosTron technology

A number of mutant strains of *C. difficile* have been developed in order to further elucidate the role of a gene of particular interest within the *C. difficile* genome. Tools have been created that disrupt the expression of the gene of interest, allowing subsequent investigation of the effects of this on the organism as a whole. The ClosTron system developed by Heap *et al*. (2007) is one such technique utilised in functional genomics studies. The ClosTron is a universal gene knockout system which employs mobile group II introns from the *ltrB* gene of *Lactococcus lactis* (L1.LtrB) to modify double-stranded DNA thereby leading to permanent disruption of the target gene (Heap *et al*., 2007; Joseph *et al*., 2018). Group II introns are designed by the modification of nucleotides within the group II intron encoding region allowing the intron to be targeted to any section of the genome. The re-targeted intron is designed using an online re-targeting algorithm on www.clostron.com (Kuehne and Minton, 2012). Mutants that contain the successfully inserted intron can be screened on erythromycin-containing media as the *C. difficile* 630 *Δ*erm strain used to generate the mutant is erythromycin resistant. The *Δ*erm strain was generated through continuous subculturing of *C. difficile* 630 daily for 30 days (Hussain *et al.* 2005). An antibiotic resistance marker (*ermB* - which codes for erythromycin resistance) is incorporated into the design of the group II introns to allow for positive selection of mutants. *ErmB* is inactivated by the insertion of a group I intron (phage *td*). When the group II intron inserts successfully into the chromosome and is transcribed, the group I intron is able to self-splice out as it is now in the correct orientation to do so, thereby reactivating the erythromycin resistance gene and allowing for positive selection of intron-containing ClosTron disruption mutants on erythromycin-supplemented media (Heap *et al*., 2007).

Examples of mutants generated by this mechanism include the *spo0A* disruption mutant created by Heap *et al*. (2007) and further investigated by Underwood *et al.* in 2009 in order to study the role of Spo0A in *C. difficile*. A *dnaK* disruption mutant was developed at Ulster University by Jain (2010) which has also been used within this project to support the investigation of the role of molecular chaperones within *C. difficile* (further discussed in Chapter 4).

### 1.13 *Clostridioides difficile* model systems

Relevant model systems for the growth and investigation of *C. difficile* are vitally important in further developing our understanding of how this pathogen behaves beyond the limited laboratory environment. Many systems have been used in recent years including animal models using mainly rodents such as hamsters, rats and mice. In addition, a number of *in vitro* models have been developed in an attempt to more closely mimic the human gut environment, such as faecal emulsions in batch and continuous culture systems (Best *et al*., 2012).

Animal models for use in the study of *C. difficile* have been in existence for over three decades. Animals used have included mice, rats, rabbits, guinea pigs and even zebrafish embryos. However, the animal most commonly used tends to be the hamster, due to the similarities found in their susceptibility to infection following antimicrobial therapy to that of humans (Sambol *et al*., 2001; Razaq *et al*., 2007). A number of studies of *C. difficile* in the hamster model helped lead to the association of *C. difficile* toxin production and antibiotic-associated colitis (Rifkin *et al*., 1978; Bartlett *et al*.,

1977; Bartlett *et al*., 1978; Borriello *et al*., 1987). Nevertheless, there are many drawbacks to the use of animals namely the obvious genetic, physical and pathophysiological differences as well as ethical issues that exist around the use of animals in research studies.

*In vitro* models that utilise human faecal matter in various forms (Gill *et al*., 2007; Gill *et al*., 2010; Brown *et al*., 2012) possess a number of advantages, including the ability to finely control these systems, which is more difficult in animal models, and the elimination of ethical issues. Continuous culture systems have been used that allow investigation over longer time periods. Studies involving continuous culture have looked at areas of *C. difficile* development such as colonisation resistance, external environmental impacts and the effect of antibiotics such as the triple stage gut model developed by McFarlane *et al*. in 1998. This system uses three anaerobic vessels controlled to maintain pH and flow rate of growth media introduction, which allows the controlled exploration of cells, toxins, spores, microflora and antimicrobial effects (McFarlane *et al*., 1998). However, there are a number of disadvantages associated with continuous culture methods as they require intensive effort and expense to maintain.

Batch culture models using faecal material within closed systems offer significant advantages such as the ability to include more replicates within the experimental design allowing result analysis to be more robust and thereby reliable. They also allow for a shorter experimental duration whilst still offering a more reflective gut environment to ordinary growth media (Best *et al*., 2012). For example, batch culture

was used in the investigation of bacteriophage (Bacteriophage ΦCD27 induced from *C. difficile* strain NCTC 12727 using mitomycin C) as a treatment option for *C. difficile* infection. It was found that there was a significant decrease in recoverable *C. difficile* after prophylactic phage administration (Meader *et al.,* 2010). Due to the success of previous studies using batch culture faecal models and the aforementioned advantages we have incorporated this method within our study.

## 1.14 AIMS OF THIS STUDY

There is a current lack of knowledge around how *C. difficile* causes disease in the real-world (i.e. outside of the artificial environment of laboratory growth media) due to limitations on the use of human and animal models. This study aims to develop an *in-vitro* model to further characterise the molecular processes underpinning pathogenesis in *C. difficile* 630. We hypothesis that understanding how *C. difficile* functions in a more biologically relevant setting will uncover mechanisms at work that cannot be determined with the use of normal laboratory media. In order to do this we will:

- Develop an *in-vitro* model to determine the growth characteristics of *C. difficile* 630 and demonstrate by RNAseq and RT-qPCR analysis significantly differentially expressed genes in order to further characterise pathogenesis in *C. difficile* strain 630.
- Further study the transcriptome of the *C. difficile* 630 mutant strains *dnaK* and *spo0A* grown in faecal water compared to the parental strain *Δerm* in order to investigate a number of known virulence-associated genes in this organism.
- Experimentally verify the presence and expression of sRNAs in *C. difficile* 630 in normal growth media and in biologically relevant growth conditions by RT-qPCR and RNA sequencing.

# CHAPTER 2: MATERIALS AND METHODS

## 2.1 Chemicals and glassware

All chemicals and reagents of Analar grade or better were purchased from Sigma-Aldrich (Poole, UK) unless stated. Brain Heart Infusion (BHI) agar and broth, and yeast extract were purchased from Oxoid (Basingstoke, UK). All molecular biology reagents were purchased from Invitrogen (Renfrewshire, UK) save for qPCR reagents, which were obtained from Roche Diagnostics (Hertfordshire, UK) and random primers, which were obtained from Promega (Southampton, UK). Lysing Matrix A tubes were from MP Biomedicals (Cambridge, UK). All apparatus was cleaned with 1% Virkon (Antec Intl Ltd., UK) and 2 % Decon (Decon Labs Ltd., UK) before use. Used glassware was left in 1% Virkon overnight and then soaked in 2% Decon for 30 minutes.

## 2.2 Growth of bacterial strains

*Clostridium difficile* 630, Δ*erm, dnaK* and *spo0A* were grown in either BHIS, BHIS/PBS or BHIS/FW media under anaerobic conditions. The anaerobic cabinet used was a Don Whitley MACS MG500 (Don Whitley Scientific Ltd, Yorkshire, UK). The anaerobic conditions within the cabinet were 80% $N_2$, 10% $CO_2$ and 10% $H_2$ at 37 °C. The growth media consisted of either 1X BHIS (brain heart infusion broth (Oxoid, Hampshire, UK) supplemented with 5g/L$^{-1}$ yeast extract (Sigma-Aldrich, Dorset, UK) and 1g/L$^{-1}$ L-cysteine (Sigma-Aldrich, Dorset, UK)) broth, 50% 2X BHIS/50% faecal water or 50% 2X BHIS/50% PBS. A starter culture was prepared by inoculating 20 ml of 50% 2X BHIS broth/50% PBS in a glass universal bottle with a single colony of freshly grown culture from BHIS agar; this was grown overnight and used to inoculate fresh media at 5% (v/v). Culture growth was recorded as attenuance

at 650 nm (D650) by transferring 800 µl or 80 µl of sample to a cuvette (depending on type of cuvette used), and measuring D650 using a Pharmacia Novaspec II spectrophotometer. The samples of culture taken for measurement were diluted 1 in 4 once an attenuance of 0.2 was reached to maintain accuracy of measurements. All equipment used including the BHIS media, $H_20$ and PBS were autoclaved before use. The faecal water was filter-sterilised through a 0.22 µm filter. Un-inoculated BHIS media and BHIS/faecal water mix were used as references. Cells were collected from all cultures at early, mid and late-log phase of growth at attenuance (D) values of approximately 0.2, 0.6, and 1.2 respectively by transferring 1 ml of each sample to an Eppendorf tube and centrifuging at 13,000 x g for 5 min at room temperature. The supernatant was discarded and the cell pellets were placed briefly in liquid nitrogen and immediately transferred to -80 °C. Strain identity was confirmed by genotyping. End-point PCR was carried out using the primers dnakF/dnakR, 4140/4620, 4140/5880, 5880/7020, EBS universal/dnaKR, spo0AF/spo0AR, EBS universal/spo0R (Table 2.2) on genomic DNA extracted from each strain.

### 2.2.1 Calculation of growth rate constant (k) and doubling time ($T_d$)

The growth rate constant (k) and doubling time ($T_d$) was calculated for each strain grown in each growth medium from attenuance collected hourly at 650 nm (D650). Exponential growth measurements from the linear part of the growth curve were used as outlined below (Neidhardt *et al*., 1990).

**Formula used to calculate growth rate constant (ƙ):**

$ƙ = (\log^{10}N - \log^{10}N0) \times 2.303/ (t\text{-}t_0)$

$N_0 = OD_{650}$nm at time 0hr

$N = OD_{650}$nm at the time of exponential growth ending

t = time of exponential growth ending (hr)

$t_0$ = time at inoculation (0hr)

**Formula used to calculate doubling time ($T_d$):**

$T_d = 0.693 / ƙ$

## 2.3 Faecal water preparation

This study was not subject to review by the Ulster University ethics committee as no intervention took place involving the volunteers who provided the samples although written consent was obtained. The samples were obtained for the express purpose of preparing growth media. Faecal samples were provided by two male donors aged between 38-42; neither had taken any antibiotics in the three months prior to collection. The samples, once collected, were stored at 4 °C for a maximum of 2 h and processed as described in Gill *et al*., 2010. This involved weighing the stool sample before mixing with ice cold 0.01 M PBS at a ratio of 1:1 w/v. This mixture was then homogenised in a Seward 600 stomacher for 2 minutes (x 2). Following homogenisation the samples were centrifuged for 2 h at 50,000 x g at 4 °C in a 70.1 Ti

rotor of a Beckman L8-M centrifuge. Supernatants were recovered and subject to filter sterilisation on ice using a 0.22 µm filter; aliquots were stored at -80 °C.

### 2.4 RNA extraction and quality control

RNA was extracted using the RNeasy® Mini kit from Qiagen (Crawley, UK). *C. difficile* 630 cell pellets containing approximately $2x10^8$ cfu were resuspended in 200 µl of lysozyme (Sigma-Aldrich, Dorset, UK) in TE buffer (15mg/ml) and 20 µl of Proteinase K (Qiagen, Crawley, UK). The pellets were incubated at room temperature for 15 min before being transferred to a lysing matrix A tube (MP Biomedicals, Cambridge, UK). The tube was placed in a FastPrep™ FP120 instrument at speed setting 5.5 for 30 s to homogenise the cells. Following this, 700 µl of buffer RLT and 7 µl of β-mercaptoethanol were added. The samples were vortexed vigorously and incubated at room temperature for 20 min. The lysate was then transferred to a 1.5 ml Eppendorf tube, making sure to avoid the beads at the bottom of the lysing matrix tube, and centrifuged at 14,000 rpm for 2 min. The supernatant was then transferred to a new 1.5 ml Eppendorf tube and 500 µl of 100% ethanol was added and the contents pipetted up and down to mix. Of this lysate 700 µl was transferred to an RNeasy Mini spin column placed in a 2 ml collection tube and centrifuged for 30 s at 14,000 rpm. The flow-through was discarded. This step was repeated for the remainder of the lysate and the flow-through again discarded. The collection tube was also discarded and replaced with a new one. To the spin column 350 µl of buffer RW1 was added and the tube inverted by hand to wash inside of the spin column. The sample was centrifuged for 30 s at 14,000 rpm, the flow-through was again discarded and the collection tube replaced with a new one. A 10 µl aliquot of DNase I stock solution (prepared by

dissolving 1500 units of DNase I (Qiagen, Crawley, UK) in 550 µl of RNase-free water) was added to 70 µl of buffer RDD and gently mixed. The total of 80 µl of the DNase I incubation mix was then added directly onto the spin column membrane and incubated for 30 min at room temperature to allow digestion of contaminating DNA. Following digestion, 350 µl of buffer RW1 was then added to the spin column and incubated for 5 min at room temperature. The column was then centrifuged at 14,000 rpm for 30 s and the flow-through discarded. The spin column was then placed into a new 2 ml collection tube and 500 µl of buffer RPE was added to the column and centrifuged at 13,000 x g for 2 min to wash the membrane. The collection tube was replaced and another 500 µl of buffer RPE was added to the column followed by further centrifugation at 13,000 x g for 2 min to ensure removal of residual ethanol from the spin column membrane; the flow-through was discarded. The column was placed inside a clean collection tube and without addition of buffer was centrifuged again for 5 min. The spin column was then placed in a new 1.5 ml collection tube and 30 µl of RNase-free water was added to the column and incubated at room temperature for 5 min, the tube was then centrifuged at 13,000 x g for 1 min in order to elute the RNA.

An in-solution DNase digestion step was carried out to ensure no genomic DNA contamination remained in the RNA sample; this was performed by mixing ≤ 87.5 µl of eluted RNA with 10 µl of DNase I stock solution and 10 µl of buffer RDD in a 1.5 ml Eppendorf tube. RNase-free water was added to the sample to make a final volume of 100 µl. The mixture was incubated for 20 min at room temperature; following this 350 µl of buffer RLT was added to the sample and pipetted up and down to mix. A

250 µl aliquot of 100% ethanol was then added and the contents of the sample transferred to an RNeasy Mini spin column placed in a 2 ml collection tube. The RNA was then purified as described previously and checked for DNA contamination by PCR followed by electrophoresis and imaging under UV light. A Nanpdrop™ 1000 spectrophotometer (Thermo Scientific) was used to quantify the amount of RNA in the samples.

**2.5 Bioanalysis of RNA samples**

The quality of the RNA was further assessed with the use of an Agilent 2100 Bioanalyzer (Agilent, USA), which uses electrophoresis to generate a ratio of the 23S and 16S ribosomal bands represented by the RNA integrity number (RIN). An algorithm produces a number between 1 and 10 to indicate the level of degradation present. Ten represents the most intact RNA sample possible, with 1 therefore representing the most degraded sample. An RNA 6000 Nano Assay kit (Agilent, USA) was used to prepare the RNA samples for analysis as per manufacturer's instructions, which included the following method. A 550 µl aliquot of RNA 6000 Nano Gel Matrix was added to a spin filter which was then centrifuged at 1,500 x g for 10 min. The filtered gel product was aliquoted into 65 µl aliquots in RNase-free microcentrifuge tubes and stored for up to 4 weeks at 4 °C. Following this, the Gel-Dye mix was prepared by initially allowing the RNA 6000 Nano dye concentrate to equilibrate to room temperature for 30 min. The RNA 6000 Nano dye concentrate was vortexed for 10 s then centrifuged at 13,000 x g for 10 s before the addition of 1 µl of dye into the 65 µl aliquot of filtered gel (previously prepared). The sample was then vortexed again for 10 s and centrifuged at 13,000 x g for 10 min at room temperature. A new RNA

chip was placed onto the chip priming station and 9 µl of gel-dye mix was loaded onto the appropriate well. The chip priming station was closed and the plunger pressed until held by the clip; the clip was released after 30 s. A 9 µl aliquot of gel-dye mix was added to the other indicated wells and the remaining gel-dye mix was discarded. A 5 µl aliquot of the RNA 6000 Nano marker was loaded onto the chip in the well labelled with a ladder symbol and in the 12 sample wells. A 6 µl aliquot was added to all unused wells. The RNA 6000 Nano ladder and the RNA samples were heated to 70 °C for 2 min before use. Following this a 1 µl aliquot of RNA 6000 Nano ladder was added to the well marked with the ladder symbol and 1 µl of RNA sample (diluted to between 50-500 ng/µl) was loaded into the sample wells. One µl of RNA 6000 Nano marker was added to each unused sample well (there must be no empty wells on the chip). The chip was placed onto the adapter of the IKA vortex mixer (Agilent, USA) and vortexed at 2,400 rpm for 1 min. The electrodes were decontaminated prior to use by loading a spare chip with 350 µl of RNaseZAP and placing this into the analyser and closing the lid for 1 min, this chip was removed and 350 µl of RNAse-free water was loaded onto another spare chip and placed inside the analyser with the lid closed for 10 s and then removed. The chip containing the RNA samples was then transferred to the bioanalyser within 5 min and run using the Agilent 2100 Bioanalyzer software as described in the manufacturer's instructions. An electropherogram was produced for each sample indicating the RIN number.

## 2.6 Transcriptome sequencing

RNA sequencing (RNAseq) and bioinformatics was carried out by Deepseq (University of Nottingham, UK). RNA samples were transported to Deepseq on dry ice, and once received, total RNA quality was again measured using the Agilent RNA 6000 Nano Kit (Agilent Technologies, USA) on the Agilent 2100 Bioanalyzer. The concentration of the RNA was calculated using the Qubit RNA BR assay kit (Life Technologies, USA). The Ribo-Zero rRNA Removal Kit (Gram-Positive Bacteria) (Illumina, Cambridge, UK) was used for rRNA depletion. NEBNext Ultra Directional RNA library prep kit for Illumina (NEB, UK) was used to prepare Illumina stranded whole transcriptome sequencing libraries. The normal procedure for preparation of Ribosome Depleted RNA was followed apart from precipitation of samples with 1 µl (20 ng/µl) glycogen and 1/10th vol. 3M sodium acetate after second strand synthesis. Pellets were washed once with 80% ethanol, then 70% ethanol and following air-drying, 58 µl of water was used to resuspend pellets. The standard Ribosome Depleted RNA protocol was recommenced for the outstanding steps, except libraries were size selected using Agencourt AMPure XP beads at a 1.5 x ratio to preserve the smaller sized fraction (~ 150bp). Barcoded multiplex libraries were generated using the NEB Next Multiplex Oligos for Illumina kit (Primer set 1) (NEB, UK). Library quality control was carried out using the Bioanalyser HS kit (Agilent Technologies, USA) and libraries were measured using qPCR (Kapa Biosystems, UK). Libraries were combined at desired concentrations, denatured and loaded for sequencing according to the manufacturer's protocol. Three runs of sequencing was performed on the Illumina MiSeq sequencing platform to create 2 x 75bp reads.

The sequencing reads were mapped onto the annotated *C. difficile* 630 reference genome (http://www.ncbi.nlm.nih.gov/nuccore/115249003) with suitable alignment software for differential gene expression analysis. The aligned files were combed for tag counts per location mapped or normalised tag counts (RPKMs) and differential gene expression analysis. The DeepSeq Filtering Pipeline for Read Mapping was utilised to filter reads with low sequencing score, as well as to reads aligned to adaptor sequences. Reads from the sequencer were quality control checked using FASTQC, then cropped and filtered for low quality bases and adaptor sequences, before being quality control checked again. The bwa mapping tool (http://bio-bwa.sourceforge.net/) was used to chart reads that passed the filter onto the reference genome in the context of known gene exon coordinates. Read alignments were saved in a BAM formatted alignment file (named *.bam), and companion BAM index file (named *.bam.bai). Both primary and unique read alignments were filtered further by their mapping quality score (MAPQ). For gene expression, MAPQ20 uniquely aligned reads were used to create counts per gene using 'htseq-count', which concludes the number of uniquely aligned reads per gene (http://www-huber.embl.de/users/anders/HTSeq/doc/count.html), which were then used as the input for the DESeq program (Anders and Huber, 2010a; Anders and Huber, 2010b). DESeq models the distribution of the counts data in each sample and then compares the distributions to determine differentially expressed genes, with significantly differentially expressed genes having an adjusted p-value $< 0.05$. A single analytical approach is used by the program and DESeq analyses the variance of the biological replicates to improve modelling of the expression of individual genes within the group.

The data generated was deposited in NCBI's Gene Expression Omnibus (Edgar *et al*., 2002) repository (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE11242), accession number: GSE112422.

### 2.7 Transcriptome data analysis

Transcriptome sequence data was obtained from DeepSeq as a summary MS Excel file containing a list of genes with cognate base mean values for BHI medium (BHI, base mean A) and faecal water medium (FW, base mean B) growth conditions, in addition to p value, p-adjusted value and the ratio of FW/BHI base mean values, sorted by p-adjusted (padj) value from low to high. Some 1218 genes had padj < 0.05. The base mean values for these genes were used to calculate log2 values for each FW/BHI ratio, from which the absolute fold-change for each gene was determined. Subsequent analysis was performed using this master list of 1218 differentially expressed (DE) genes with pdaj < 0.05 and FC > 1.5. The NCBI *C. difficile* 630 genome (http://www.ncbi.nlm.nih.gov/nuccore/115249003) was used for addition of the *C. difficile* 630 locus annotations (Sebaihia *et al*., 2006; Monot *et al*., 2011), in addition to protein name and Subtilist functional category (Moszer *et al*., 2002; Lechat *et al*., 2008). This process was carried out as described previously (Ternan *et al*., 2012; Ternan *et al*., 2014) using the NCBI CDD database, BioCyc pathway tools and MetaCyc metabolic pathways software (Marchler-Bauer *et al*., 2015; Caspi *et al*., 2016; Karp *et al*., 2016) in conjunction with literature searching to determine a functional role/category and to ascertain predicted co-regulated genes and operon arrangements.

## 2.8 Reverse transcription

cDNA was prepared from the extracted RNA samples by adding 50 ng of random hexamer primer (Promega, WI, USA) to a 0.2 ml Eppendorf tube, to this 500 ng of RNA sample was added. The volume was made to 12 µl with RNase-free water and the mixture was incubated at 70 °C for 10 min. To each tube 4 µl of 5 X First strand buffer, 2 µl of 0.1 M DTT, 1 µl of 10 mM dNTP mix and 1 µl of RNaseOUT™ Ribonuclease Inhibitor (Invitrogen, Paisley, UK) was added. The tubes were incubated at 42 °C for 2 min. Following incubation, 200 U of SuperScript™ Reverse Transcriptase (1 µl) was added followed by a further incubation step at 42 °C for 50 min. A final incubation at 70 °C for 15 min was completed in order to inactivate the SuperScript™ enzyme. An "RT negative" control was also prepared alongside the other samples, which contained all of the reaction solutions apart from the reverse transcriptase enzyme.

## 2.9 Primer design and validation

The primers used in this thesis (unless otherwise stated) were designed using Primer3Plus software version 2 (Untergasser and Nijveen, 2007). In order to save on time, money and the chance of variability the melting temperatures were kept as similar as possible as this would allow more than one gene to be analysed on a plate at once. Other parameters taken into consideration included the amplicon size, which for qPCR purposes was less than 250 bp, the GC content was kept to between 20-80% and the primer size between 18-27 bp. Primers were reconstituted in nuclease-free water to a concentration of 30 pmol/µl and stored at -80°C in 100 µl aliquots. End-point PCR was used to validate the primers on genomic DNA extracted from *C.*

*difficile* 630. Only those primer pairs that produced a clear bright band on gel electrophoresis imaged under UV light were used in subsequent experiments.

### 2.10 Polymerase chain reaction (PCR)

PCR was carried out using a TGradient Thermocycler (Biometra, Goetigen, Germany). All PCR reagents were purchased from Invitrogen (Paisley, UK). The PCR master mix was made up of 100 µl of 10 X PCR buffer, 30 µl of 1.5 mM $MgCl_2$, 2 µl of each dNTP from a 100 mM stock and made up to 1 ml with 827 µl of millipore water; 96.5 µl of this master mix was added to a 0.2 ml PCR tube, to this 1.5 U of Taq polymerase, 1 mM of forward and 1 mM of reverse primer (30 pmol/µl) and 1 µl of template was added to make up a total 100 µl PCR reaction. The PCR cycling conditions used were as follows: an initial denaturation stage for 5 min at 95 °C followed by 30 cycles of 95 °C for 30 s, an annealing stage at 57 °C for 30 s and extension at 72 °C for 30 s with a final extension at 72 °C for 5 min.

### 2.11 DNA extraction – Fast DNA Spin Kit method

DNA was extracted using the Fast DNA Spin Kit (MP Biomedicals, Leicester, UK) from *C. difficile* 630 pellets containing approximately 4 x $10^8$ cells. The cell pellet was resuspended in CLS-TC buffer by vortexing for 10 s. The suspension was then transferred to a lysing matrix A tube (MP Biomedicals, Leicester, UK) and a FastPrep®FP120 cell disrupter (Qbiogene, France) was used to homogenize the solution for 40 s at a speed setting of 6.0. Following this, the sample was centrifuged for 10 min at 13,000 x g in an Eppendorf 5418R Centrifuge. The supernatant was

carefully removed and added to a 15 ml polypropylene tube with an equal volume of binding matrix. Gentle agitation was performed for 5 min on a Luckham 802 Suspension Mixer rotator at speed setting 7. A 700 µl aliquot of the suspension was transferred to a SPINFilter™ placed inside a catch tube and centrifuged for 1 min at 13,000 × g. The catch tube was emptied and this step was repeated until all sample had been transferred through the matrix by centrifugation. The binding matrix remaining on top of the SPINFilter was resuspended in 500 µl of SEWS-M wash solution by pipetting up and down several times. The SPINFilter was placed inside a clean catch tube and centrifuged for 1 min at 13,000 x g. The catch tube was emptied and the SPINFilter was centrifuged again for a further 2 min at 13,000 x g in an effort to remove any remaining ethanol from the wash step. The SPINFilter was placed inside a sterile 1.5 ml Eppendorf tube and 100 µl of DES (DNase-free water) was used to gently resuspend the binding matrix. This was followed by incubation at 55 °C for 5 min on a Grant microtube block thermostat (UK); the Eppendorf containing the SPINFilter was centrifuged for 1 min at 13,000 × g. Following this, the SPINFilter was discarded and the extracted DNA was quantified using the Nanodrop™ 1000 Spectrophotometer (Thermo Fisher Scientific, Loughborough, UK). A 5 µl aliquot of the extracted DNA was electrophoresed on a 1.5% agarose gel to visualise and the DNA was stored in 10 µl aliquots at -20 °C.

### 2.12 DNA extraction – Chelex method

*C. difficile* 630 pellets containing approximately 4 x $10^8$ cells were washed by resuspension in 1 ml of ice-cold PBS followed by centrifugation at 13,000 x g for 10 min at 4 °C in an Eppendorf 5418R Microcentrifuge (Stevenage, UK). The resultant

supernatant was discarded and the cell pellet was again resuspended in 1 ml of ice-cold PBS by vortexing for 5-10 s. Following this, the suspension was centrifuged again at 13,000 x g for 10 min at 4 °C and the supernatant was discarded; this step was performed a further time. A 5% w/v Chelex (Sigma-Aldrich, UK) solution was prepared in PBS and the cell pellet was resuspended in 300 µl of this solution. A Grant microtube block thermostat (UK) was used to incubate the solution at 100 °C for 15 min. Centrifugation was carried out at 13,000 x g for 10 min at 4 °C to separate the now in-solution DNA from the Chelex resin. The supernatant was transferred to a sterile Eppendorf and stored at -20 °C.

### 2.13 Gel electrophoresis

A 1.5% agarose gel was prepared using molecular biology grade agarose (Qbiogene, Quebec, Canada) and 1X TBE buffer (Invitrogen, Paisley, UK). Ethidium Bromide (Sigma-Aldrich, Dorset, UK) was added to the gel (4 µl/100 ml gel) to enable visualisation of the nucleic acid under UV light. A sample of 8 µl was mixed with 2 µl of loading dye (Sigma-Aldrich, Dorset, UK) and loaded into the gel. Four µl of 100 base pair ladder (Invitrogen, Paisley, UK) was loaded as a reference to allow the product size to be estimated. The gel was electrophoresed at 100 v for 20 min and imaged using the Gel Doc™ EZ imaging system (Bio-Rad, Hertfordshire, UK).

### 2.14 RT-qPCR

All quantitative PCR (qPCR) was carried out using the LightCycler480 instrument (Roche Diagnostics, UK). A standard curve was prepared by creating a 5-fold serial

dilution (1, 1:5, 1:25, 1:125, 1:625, 1:3125, and 1:15125) of the pooled cDNA samples with nuclease-free water. Each reaction well contained 5 µl of Master SYBER Green I (Roche Diagnostics, UK), 1 µl of forward primer, 1 µl of reverse primer (at a concentration of 10 µM), 2 µl of nuclease free $H_2O$, and 1 µl of template. A 10 µl aliquot of reaction mix was added to each well on the 96-well plate in triplicate. A 1 in 10 dilution of each cDNA sample was used for the target run templates. The programme conditions used for the qPCR were: an initial denaturation stage of 95 °C for 5 min followed by 40 cycles of 95 °C for 10 s, 57 °C for 10 s and 72 °C for 10 s. A melting curve was performed once the programme had finished and a report was generated to allow for subsequent analysis of the data. The Minimum Information for publication of Quantitative real-time PCR Experiments (MIQE) guidelines published by Bustin *et al*. in 2009 were followed when designing and carrying out all of the RT-qPCR experiments described in this thesis.

### 2.15 RT-qPCR data analysis

The LC480 instrument software (version 1.5) was used in combination with SYBR Green I detection (Roche Diagnostics) to generate a $C_q$ value for each sample using the 'second derivative maximum' method. Each $C_q$ value corresponds to the point at which the level of fluorescence in the sample crosses the threshold of background fluorescence, which is inversely proportional to the amount of nucleic acid present in the sample. This $C_q$ value was used to analyse the data to determine expression levels. $C_q$ values were transferred to Excel and initially the arithmetic mean of the technical replicates was determined. This value was then log transformed to relative quantities (RQ) using the information gained from the standard curves performed for each primer

pair i.e. the mean of the technical replicates minus the y-intercept value was calculated and then divided by the slope of the standard curve, this number was then back-transformed to produce the RQ value. This method ensured the PCR efficiencies were calculated accurately for each gene and no assumptions were made. All target gene RQs were normalised against the geometric mean of three reference gene RQs by dividing the former by the later to produce the normalised relative quantity (NRQ) value. The untreated control samples (BHIS/PBS) were used as a calibrator and corrected to 1 with all experimental samples (BHIS/FW) expressed as a relative expression ratio to the calibrator for each gene. It is the geometric mean of the biological replicate calibrator normalised relative quantity (CNRQ) values that are reported.

### 2.16 PCR clean-up

The PCR products were cleaned up for sequencing using the Wizard SV Gel and PCR Clean-up System (Promega, USA). Firstly, an equal volume of Membrane Binding Solution was added to the PCR product and then transferred to an SV minicoloumn placed inside a collection tube. This solution was incubated at room temperature for 1 min followed by centrifugation at 13,000 x g for 1 min. The flow-through was discarded at this point and the minicolumn reinserted into the collection tube. The product was then washed by adding 700 µl of Membrane Wash Solution to the minicolumn and centrifuging at 13,000 x g for 1 min, the flow-through was again discarded and the minicolumn reinserted into the collection tube. This step was repeated with 500 µl of Membrane Wash Solution, followed by centrifugation at 13,000 x g for 5 min. The collection tube was emptied followed by re-centrifugation

of the column assembly for 1 min with lid open to allow evaporation of any residual ethanol. The minicolumn was then transferred to a clean 1.5 ml microcentrifuge tube and 50 µl of nuclease-free water was added to the minicolumn. The column assembly was incubated at room temp for 1 min followed by centrifugation at 13,000 x g for 1 min to elute the PCR product. The mincolumn was then discarded and the elution was stored at -20 °C until needed.

**2.17 DNA sequencing**

The samples were prepared for sequencing using BigDye Terminator version 3.1 Cycle sequencing kit using the following protocol: For each sequencing reaction, the following components were added to a sterile 0.2 ml tube: 1 µl of ready-to-use ready reaction mix, 2 µl 5X sequencing buffer, 1 µl of primer at a concentration of 3.2 pmol/ µl (forward and reverse primers were used individually in separate reactions), and 1-5 µl of purified PCR product depending on amplicon size. The reaction was made up to a final volume of 20 µl with sterile deionised water. The reaction mixture was then PCR amplified on a Biometra Tpersonal Thermocycler using the following parameters: an initial denaturation at 96 °C for 1 min followed by 25 cycles of denaturation at 96 °C for 10 s, an annealing stage at 50 °C for 5 s and extension at 60 °C for 4 min.

**2.18 Purification of sequencing products by gel filtration**

The products were purified using Performa DTR Gel Filtration columns (Edge Bio, Maryland, USA) by the following method: The liquid was removed from Performa

DTR Gel Filtration column by centrifuging at 900 rpm for 3 min. The column was then transferred to a labelled 1.5 ml tube and the sequencing reaction was added drop-wise avoiding contact with the gel. The sample was incubated at room temperature for 1 min to allow absorption of the BigDye by gel. The sequencing product was then collected by centrifuging for 1 min at 900 rpm. The elute was retained and the column discarded. A master mix was prepared for precipitation of the DNA containing 64 µl of 95% ethanol and 16 µl of $dH_2O$. To each sample 80 µl of the master mix was added. The samples were incubated overnight at -20 °C and then centrifuged at 13,000 x g for 20 min. The supernatant was aspirated using suction and 100 µl of 80% ethanol was added to each sample followed by incubation at room temperature for 5 min. The samples were centrifuged at 13,000 x g for 10 min at 4 °C and the supernatants were aspirated before being placed on a heat block at 65 °C for 5 min to allow evaporation of residual ethanol. All samples were then prepared for sequencing on the ABI PRISM 3100 genetic analyser (Applied Biosystems, Foster City, CA) by adding 20 µl of formamide to each reaction mixture to stabilise the DNA strands. The samples were incubated at 95 °C for 5 min and then transferred directly onto ice for 3 min followed by brief centrifugation. The samples were then loaded onto a 96 well plate before being placed into the analyser. The DNA sequences were analysed using Chromas software.

### 2.19 Endospore staining

Liquid bacterial culture was collected at each timepoint and 40 µl was pipetted onto a labelled clean glass slide and allowed to air dry. The bacterial smears were then heat fixed by passing through a flame several times. A Bibby HC502 hot plate was set to 150 °C and a beaker of water placed on top. Once the water was boiling the slides were

placed on top of the beaker and covered with 1.5% malachite green stain for 5 min. Malachite green stain was continually applied throughout this time in order to prevent drying of the sample. Following this, excess malachite green stain was removed and the slides were rinsed with deionised water. The slides were then flooded with a counterstain of 1% safranin for 30 s. Slides were again rinsed with deionised water and blotted with bibulous paper and viewed under 100 X oil immersion using a Nikon Eclipse E400 microscope equipped with DS-5M-L1 Digital Sight Camera and NIS-Elements Basic Research 3.2 software. Spores were counted from 10 separate fields of view per slide.

**2.20 Gram's staining**

Liquid bacterial culture was collected at each timepoint and 40 µl was pipetted onto a labelled clean glass slide and allowed to air dry. The bacterial smears were then heat fixed by passing through a flame several times. A primary stain of 0.1% crystal violet was applied in excess for 30 s then washed away with deionised water. Following this, Gram's iodine was applied for 30 s and then washed away with deionised water. The slides were washed with 95% alcohol for 10 s and then rinsed with deionised water. A counterstain of 1% carbol fuschin was applied for 1 min. Slides were again rinsed with deionised water and blotted with bibulous paper and viewed under 100 X oil immersion using a Nikon Eclipse E400 microscope equipped with DS-5M-L1 Digital Sight Camera and NIS-Elements Basic Research 3.2 software. 100 cells were measured per slide.

**Table 2.1:** Primers used in this study.

| Gene | Locus | Description | Primer | Sequence (5′→3′) | Product size(bp) | Annealing temp (°C) | Reference |
|---|---|---|---|---|---|---|---|
| *rho* | CD3487 | Transcription termination factor Rho | rho-F | CATCAAGCAATAAATCATCTC | 153 | 57 | Metcalf *et al*, 2010 |
| | | | rho-R | CTGGTTCTAGGATGGATGATG | | | |
| *rrs* | rrs | 16S ribosomal RNA | rrs-F | GGGAGACTTGAGTGCAGGAG | 120 | 57 | Denève *et al*, 2008 |
| | | | rrs-R | GTGCCTCAGCGTCAGTTACAGT | | | |
| *gyrA* | CD0006 | Gyrase subunit A | gyrA-F | CTCGTATTGTTGGGGACGTT | 146 | 57 | Denève *et al*, 2008 |
| | | | gyrA-R | ATCCCCATCAACAGAACCAA | | | |
| *adk* | CD0091 | Adenylate kinase | adk-F | GTGTATGTGATGTATGCCAAG | 196 | 57 | Metcalf *et al,* 2010 |
| | | | adk-R | CCTAAGGCTGCGACAATATC | | | |
| *rpsJ* | CD0072 | 30S ribosomal protein S10 | rpsJ-F | GATCACAAGTTTCAGGACCTG | 151 | 57 | Metcalf *et al*, 2010 |
| | | | rpsJ-R | GTCTTAGGTGTTGGATTAGC | | | |
| *groES* | CD0193 | 10 kDa chaperone | groES-F | AGTTTTACCAGGAGCAGCTAAAG | 116 | 57 | This study |
| | | | groES-R | CCTTATCTCCCACTGTCAATTCC | | | |
| *groEL* | CD0194 | 60 kDa chaperone | groEL-F | TTGCTGGAGGAGTAGCTGTTG | 143 | 57 | This study |
| | | | groEL-R | AAAAGCAGTTCCTCCACCAG | | | |
| *fliC* | CD0239 | Flagellin subunit | fliC-F | TGATGATGCTGCTGGACTTG | 119 | 57 | This study |
| | | | fliC-R | ACGAACCTTCTGCTGTTTGTAC | | | |
| *fliD* | CD0237 | Flagellar cap protein | fliD-F | AGCTGGACAAATTGCCAGTG | 114 | 57 | This study |
| | | | fliD-R | CCTTGGTCATCAGTTACATCAGC | | | |
| *tcdA* | CD0663 | Toxin A | tcdA-F | AGCTTTCGCTTTAGGCAGTG | 129 | 57 | This study |
| | | | tcdA-R | ATGGCTGGGTTAAGGTGTTG | | | |
| *spoVB* | CD3498 | Stage 5 sporulation protein B | spoVB-F | ATTCAGGGAATGGGAAAACC | 161 | 57 | This study |
| | | | spoVB-R | TTAATCATGGCTGCCACAAA | | | |
| *cspC* | CD2246 | Subtilisin-like serine germination | cspc-F | AGTGGTGCAGGAAATCAAGG | 155 | 57 | This study |
| | | | cspc-R | CTGTGCTCCCACTTTATCTGG | | | |
| *spoIII AC* | CD1194 | Stage III sporulation protein AC | spoIIIAC-F | ATCATTGATATTAAAGGTTGCAGGT | 150 | 57 | This study |
| | | | spoIIIAC-R | TCACTTATTTCTGTTATTACCATACCC | | | |
| *dnaK (1)* | CD2461 | Chaperone protein | dnaK-F | GATGCAGGACTTTCTACAGGTG | 144 | 57 | This study |
| | | | dnaK-R | TGCTGCAACACACTCATCTG | | | |
| *dnaK (2)* | CD2461 | Chaperone protein | dnaK-F | CTACAGCTGGTAACAATAGATTAGGT | 210 | 57 | Jain *et al*., 2010 |
| | | | dnaK-R | CTGTAGCAGTTATGAAAGGTAAGTT | | | |
| *dnaJ* | CD2460 | Chaperone protein | dnaJ-F | ATTTGGCGGTCAAGGCTTTG | 112 | 57 | This study |
| | | | dnaJ-R | ACGTTGAGGCCCTCTTCTTC | | | |
| *tig* | CD3306 | Trigger factor | tig-F | GCTGAAGGAATTGAAGCAACAG | 112 | 57 | This study |
| | | | tig-R | TGTCGGCATCTCTTAAAGAAGC | | | |
| *grpE* | CD2462 | Heat shock protein | grpE-F | GACTTCAAGCCGAATATGCAAAC | 138 | 57 | This study |
| | | | grpE-R | CACATGCATCTAAGGCTCTTTCC | | | |
| *tcdB* | CD0660 | Toxin B | tcdB-F | GAGAAGGGCATAATGAGAACGG | 137 | 57 | This study |
| | | | tcdB-R | TGCATGACACCATCTTCACC | | | |
| *tpi (1)* | CD3172 | Triosphosphate isomerase | tpi-F | AAAGAAGCTACTAAGGGTACAAA | 227 | 57 | Lemèe *et al.*, 2004 |
| | | | tpi-R | CATAATATTGGGTCTATTCCTAC | | | |
| *tpi (2)* | CD3172 | Triosphosphate isomerase | tpi-F | GCAGGAAACTGGAAAATGCATAA | 488 | 55 | Lemèe *et al.*, 2004 |
| | | | tpi-R | CAGATTGGCTCATATGCAACAAC | | | |
| *rpoA* | CD0098 | RNA polymerase subunit A | rpo0A-F | GGATGATATGATGAAGGTTAGAAACCT | 81 | 57 | O'Connor *et al.*, 2006 |
| | | | rpo0A-R | CCCAATCCAAGTTCTTCTAGTTTTTG | | | |
| *gluD* | CD0179 | Glutamate dehydrogenase | gluD-F | ATGCAGTAGGGCCAACAAAA | 135 | 57 | Denève *et al*, 2008 |
| | | | gluD-R | TTCCACCTTTACCTCCACCA | | | |

**Table 2.2:** Primers used in genotyping of *C. difficile* strains. See Chapter 4, Table 4.2 for amplicon expected.

| **Primer** | **Sequence (5′→3′)** | **Target** | **Annealing temp (°C)** | **Reference** |
|---|---|---|---|---|
| 4140-F | TAAGAGTGTGTTGATAGTGC | Erm leader peptide and erm1 (B) | 57 | Hussain *et al.,* 2005 |
| 4620-R | TCAATAGAGCGTTACCTGTTTAC | Erm leader peptide and erm1 (B) | 57 | Hussain *et al.,* 2005 |
| 5880-F | CGACTATGTTTGCTTACTTGAC | ORF298 and erm2 (B) | 57 | Hussain *et al*., 2005 |
| 5880-R | GTCAAGTAAGCAAACATAGTCG | Erm leader peptide, erm1 (B), ORF3 and ORF 298 | 57 | Hussain *et al*., 2005 |
| 7020-R | TCAATAGAGCGTTACCTGTTTAC | ORF298 and erm2 (B) | 57 | Hussain *et al*., 2005 |
| EBS Universal | CGAAATTAGAAACTTGCGTTCAGTAAA | Screening mutant | 57 | Heap *et al*., 2007 |
| *dnaK*-F | CTACAGCTGGTAACAACAATAGATTAG | *dnaK* gene | 57 | Jain *et al*., 2017 |
| *dnaK*-R | CTGTAGCAGTTATGAAAGGTAAGTT | *dnaK* gene | 57 | Jain *et al*., 2017 |
| *spo0A*-F | TTAGGCATAGCTAAGGATGGA | *spo0A* gene | 57 | Heap *et al*., 2007 |
| *spo0A*-R | TTAGGCATAGCTAAGGATGGAA | *spo0A* gene | 57 | Heap *et al*., 2007 |

**Table 2.3:** sRNA primers used in this study. All primers were designed to perform optimally with an annealing temperature of 57 °C.

| sRNA ID | Description | Primer | Sequence (5′→3′) | Amplicon (bp) |
|---|---|---|---|---|
| scdf3688.1 | SsrA | scdf3688.1-F | ACGTGGGTTTGGAACTTGAG | 147 |
| | | scdf3688.1-R | CCAGTCTTAGTCGGCAGGAG | |
| scdf3148.1 | RNaseP_bact_a | scdf3148.1-F | CCATAGAGCAGGGTGCTAGG | 108 |
| | | scdf3148.1-R | AGCTCTTACCTCGCCTTTCC | |
| scdf21.1 | ffs | scdf21.1-F | GTGCCCTGTAACCTGCAATC | 156 |
| | | scdf21.1-R | CACTTACCGCTGCTTCCTTC | |
| scdf2295.1 | T-box | scdf2295.1-F | AGAGAGCTGAGGATGGTGGA | 200 |
| | | scdf2295.1-R | CGCGTTGCCACCTAAATTAT | |
| scdf2056.1 | T-box | scdf2056.1-F | GTTTCCGATTGGTGAGAGGA | 194 |
| | | scdf2056.1-R | CTCGTGGTTCCACCTCTGTT | |
| scdf3815.1 | T-box | scdf3815.1-F | GAGAGGAAAATTCACTGGCTGT | 209 |
| | | scdf3815.1-R | AAGCTCGCGTTACCACCTTA | |
| scdf1811.1 | T-box | scdf1811.1-F | CCACAGAGAACCAATGTTGC | 196 |
| | | scdf1811.1-R | CGCGGTACCACCTTATTTGT | |
| scdf3115.1 | T-box | scdf3115.1-F | CGAGTTGGTGATGGTGTGAG | 212 |
| | | scdf3115.1-R | AAAAGGGACGAAAGGTTGCT | |
| scdf1763.1 | T-box | scdf1763.1-F | TCCAGTTTGGTGAAATGGAAT | 179 |
| | | scdf1763.1-R | CGTGTTACCACCTTAGTTCATCA | |
| scdf1153.1 | T-box | scdf1153.1-F | CAGAGAGCTAAAGGATGGTGTG | 189 |
| | | scdf1153.1-R | GCTCGCGGTACCACTCTATT | |
| CD630_s0030 | T-box (Pro) | CD630_s0030-F | ACAGAGAGAAAACGGTGGTGA | 145 |
| | | CD630_s0030-R | TCCGCGGTACCACTCTAGTT | |
| CD630_n00170 | ncRNA IGR | CD630_n00170-F | TTAAATAGTGGCAACAATTGAGGA | 60 |
| | | CD630_n00170-R | AATACCCTATACAAATAGTGGCAA | |
| CD630_n00330 | ncRNA IGR | CD630_n00330-F | CTTTTGGACAGCTCAAGTAGTGA | 154 |
| | | CD630_n00330-R | ACTCTTATCCACCCTAACCCT | |
| CD630_s0360 | FMN | CD630_s0360-F | ACTTTACTGTCGGCTTTGGA | 92 |
| | | CD630_s0360-R | GGGGTAGGGTGAAATTCCCAAT | |
| CD630_s0410 | Group I intron | CD630_s0410-F | CTACGCACACTCGGGACTTT | 126 |
| | | CD630_s0410-R | AGCTTTAGGGGTTTCCAGCA | |
| CD630_s0490 | T-box (Asn) | CD630_s0490-F | AACGGTGCCTTCCGTCTAAG | 109 |
| | | CD630_s0490-R | AGCGAGTTGAGGATGGTGAA | |

# CHAPTER 3: TRANSCRIPTOMIC ANALYSIS OF *CLOSTRIDIOIDES DIFFICILE 630* GROWN IN A FAECAL WATER MEDIUM

## 3.1 INTRODUCTION

*Clostridioides difficile* is a Gram-positive spore forming anaerobe and widely regarded as a chief cause of nosocomial infections (Rupnik *et al*., 2009; Burke and Lamont, 2014). The development of disease from *C. difficile* infection is most likely to occur when the patient is receiving a course of broad spectrum antibiotics for an unrelated bacterial infection, this leads to a disturbance in the patient's natural microflora permitting growth of the pathogen and disease development (Jarrad *et al*., 2015; Czepiel *et al*., 2019). Although it is well known that the gut microflora is affected by antibiotic consumption which can more readily lead to the establishment of *C. difficile* infection (CDI), how this process actually occurs is only just beginning to be elucidated. Changes in the colonic environment that can lead to the development of CDI have been linked to both dietary and microflora compositional fluctuations (Moore *et al*., 2015; Samarkos *et al*., 2018). There have been a number of model systems described in the literature for the study of *C. difficile* pathogenesis, however those which involve human participants are limited to prospective or retrospective testing for the interpretation of *C. difficile* strain types (Best *et al*., 2012). It could therefore be said that there are few studies in the literature currently that have developed experimental systems to study *C. difficile* virulence with direct relevance to the human gut environment. Faecal water (FW) however, which is the aqueous component of human faeces, offers a means of studying the relationship between the contents of the colon and the presence of *C. difficile* and thereby link the presence of this pathogen with gut health (Rafter *et al*., 1987; Eid *et al*., 2015; Pearson *et al*., 2009; Daniela *et al*., 2014; Verbeke *et al*., 2015; Windey *et al*., 2015). Due to the availability of unbound, soluble components such as, fatty acids, short-chain fatty acids, amino acid residues and bile acids etc. FW has been utilised as a biologically-relevant

challenge agent in a variety of gut health investigations (Boyd *et al*., 2006; Monleón *et al*., 2009; Gill *et al*., 2010; Brown *et al*., 2012; Claesson *et al*., 2012; Nowak *et al*., 2014). These components are likely to play an important role in the composition and function of the human gut microbiome therefore we have combined FW with ordinary BHIS growth media as a way of simulating the human colonic setting and allowing subsequent biologically appropriate modelling of *C. difficile.*

Using RNA sequencing (RNAseq) we will harness next-generation sequencing technology to assess the presence and quantity of RNA in our samples at the point of cell harvest (Wang *et al*., 2009; Chu and Corey, 2012). RNAseq has been developed to allow analysis of the cellular transcriptome of a given sample with comparison to a control. RNAseq can identify total RNA as well as mRNA, small RNA, and tRNA transcripts (Maher *et al*., 2009; Ingolia *et al*., 2012; Stark *et al*., 2019). Before the introduction of RNAseq gene expression studies relied heavily on hybridization-based microarrays which were prone to issues such as cross-hybridization artifacts, unreliable detection of low abundance transcripts as well as the requirement for prior knowledge of the sequence of interest (Draghici *et al*., 2006; Kukurba and Montgomery, 2015). In addition to RNAseq we will use RT-qPCR to assess a selection of genes in order to strengthen the reliability of our data.

## 3.2 AIMS AND OBJECTIVES

This study aims to utilise a unique *in vitro* technique to demonstrate by RNAseq and RT-qPCR analysis which genes are significantly differentially expressed in the faecal water media in order to further characterise pathogenesis in *C. difficile* 630.

The main objectives for this chapter are:

- Design an *in vitro* model for investigation of the growth of *C. difficile* 630 in an environment more analogous to the human intestine.
- Extract high quality and quantity RNA to ensure the reliability of all downstream applications.
- Use RNAseq to establish the transcript status of the cells at the point of cell collection (late-log phase).
- Confirm RNAseq results with the use of carefully designed and executed RT-qPCR analysis.

## 3.3 RESULTS

### 3.3.1 Growth of *Clostridioides difficile* 630

In order to study the growth of *C. difficile* 630 in a more biologically replicative environment, the organism was grown in three separate mediums i.e. BHIS, BHIS/faecal water (FW) (50/50 v/v), BHIS/PBS (50/50 v/v). Attenuance (650 nm) was measured in triplicate over a time period of seven hours. Statistical analysis showed that the presence of FW in the media was not detrimental to growth of *C. difficile* 630 over this period of time (Figure 3.1).

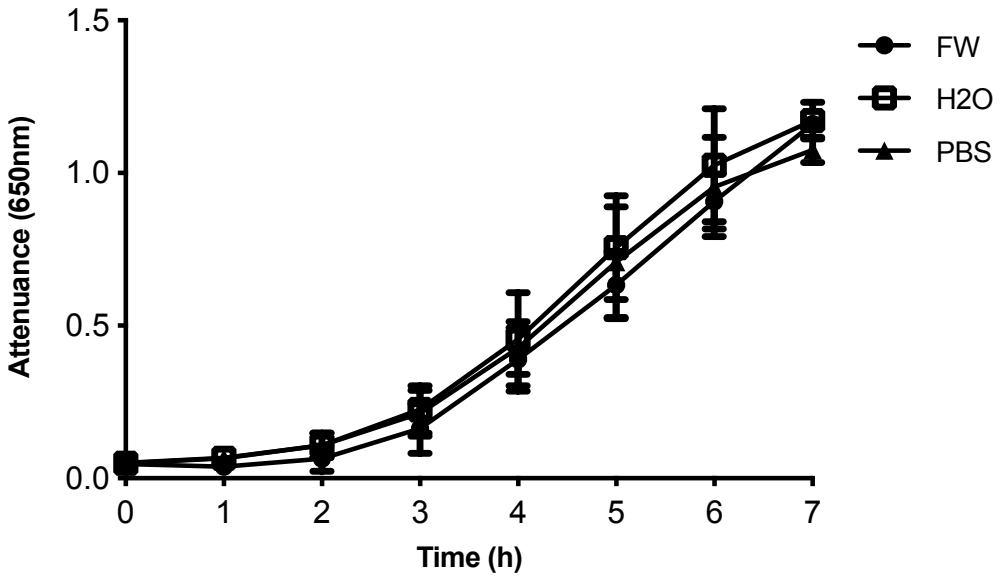



**Figure 3.1:** Growth (D650 nm) of *Clostridioides difficile* 630 in BHIS containing 50% faecal water, BHIS containing 50% PBS, and standard BHIS made with deionized $H_2O$. Data presented are means of three independent biological replicates and error bars represent the standard deviation of the mean.

### 3.3.2 RNA extraction and QC

RNA was extracted from cell pellets of *C. difficile* 630 grown at 37 °C as described above. The RNA extraction method used (described in Chapter 2) was modified slightly from the RNeasy® Mini kit (Qiagen, UK) manufacturer's instructions to increase RNA quality and yield. Changes to the protocol included incubation of the sample in lysing buffer for 20 minutes to help break open the Gram-positive cells. An extra spin step was performed (without the addition of buffer) in order to remove residual ethanol and buffers from the final RNA elute. The resultant RNA was visualised by gel electrophoresis (Figure 3.3) and the yield and quality determined through Nanodrop spectrophotometry (Table 3.1). PCR was carried out using the extracted RNA as template with *tpi* primers to ensure no gDNA contamination was present (data not shown) - this is an especially important step before reverse transcription of the RNA as the presence of genomic DNA in RT samples will lead to false positive results. The RNA was further analysed using the Agilent 2100 Bioanalyzer. The Bioanalysis showed that the RNA was of high quality (RIN > 9.0) and free of degradation and residual contaminants such as buffers and salts from the extraction process (Table 3.1 and Figure 3.2). This was a fundamental step prior to RNA sequencing to ensure good quality reads were produced.

**Table 3.1:** Bioanalysis of RNA extraction from *Clostridioides difficile* 630 grown in BHIS containing 50% faecal water (FW), BHIS containing 50% PBS and standard BHIS made with deionized $H_2O$. Three biological replicates were assessed for each growth condition.

| RNA Sample | Concentration (ng/µl) | A260/280 | A260/230 | RIN |
|---|---|---|---|---|
| **FW Rep 1** | 1966.8 | 2.18 | 2.50 | 9.1 |
| **FW Rep 2** | 2213.1 | 2.18 | 2.48 | 9.3 |
| **FW Rep 3** | 2254.7 | 2.17 | 2.49 | 8.6 |
| **PBS Rep 1** | 2097.4 | 2.18 | 2.54 | 9.5 |
| **PBS Rep 2** | 1835.8 | 2.18 | 2.38 | 9.3 |
| **PBS Rep 3** | 1888.3 | 2.18 | 2.51 | 9.6 |
| **H2O Rep 1** | 1998.5 | 2.18 | 2.43 | 9.3 |
| **H2O Rep 2** | 1504.8 | 2.18 | 2.48 | 9.3 |
| **H2O Rep 3** | 2047.8 | 2.18 | 2.44 | 9.4 |

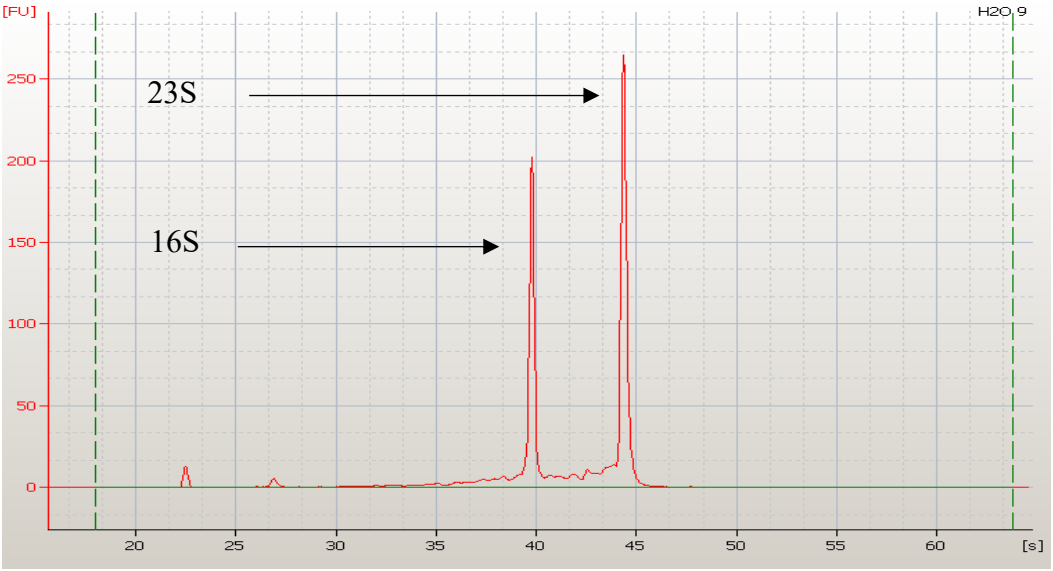



**Figure 3.2a:** Representative chromatogram generated from the bioanalysis of RNA extracted from *Clostridioides difficile* 630 grown to late-log phase at 37 °C in BHIS broth. RIN of 9.6.

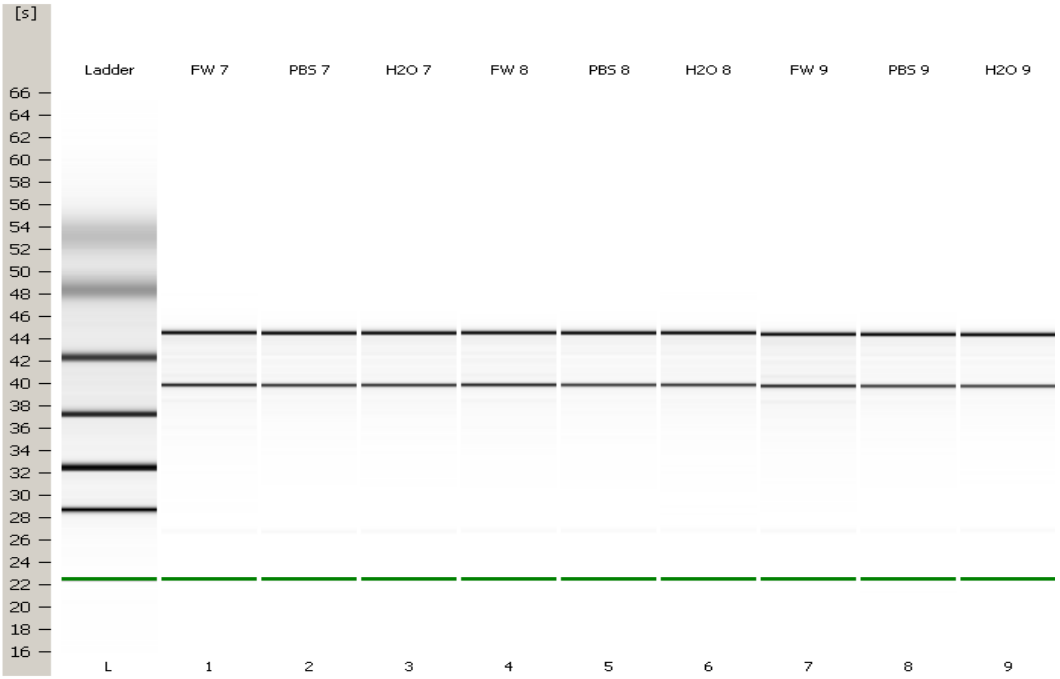



**Figure 3.2b:** Representative gel image generated from the bioanalysis of RNA extracted from *Clostridioides difficile* 630 grown to late-log phase at 37 °C in BHIS broth.

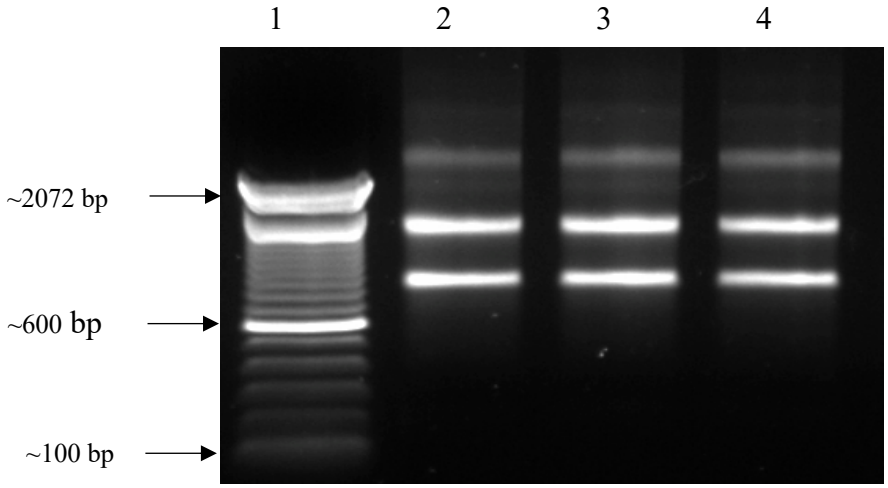



**Figure 3.3:** Gel electrophoresis image of RNA extracted from *Clostridioides difficile* 630. **Lane 1:** 100 bp molecular weight marker; **Lanes 2-4:** RNA extracted from *Clostridioides difficile* 630 grown to mid-log phase at 37 °C in BHIS broth.

### 3.3.3 cDNA synthesis and validation

cDNA was synthesised successfully from the extracted RNA (Figure 3.4), in order to determine the presence of cDNA and absence of gDNA contamination within the samples PCR was performed for each reaction. Figure 3.4 shows that an amplicon corresponding to the *tpi* primer pair is present. The negative RT controls, which were set up alongside the cDNA samples and contain all of the reagents apart from the reverse transcriptase enzyme, show that there is no contamination present.

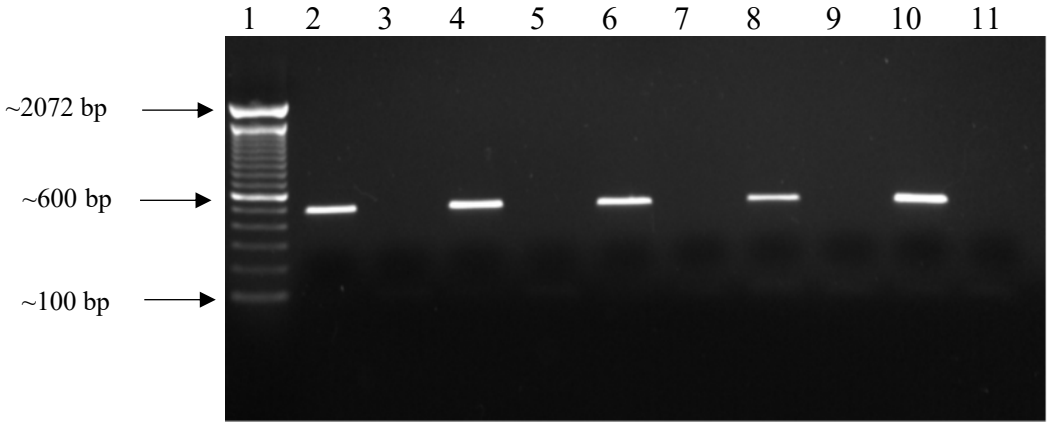



**Figure 3.4:** Gel electrophoresis image of cDNA synthesised using reverse transcriptase and amplified by PCR using *tpi* primers with a product size of 488 bp. Each RT product has a control containing no reverse transcriptase enzyme. **(A) Lane 1:** 100 bp molecular weight marker. **Lane 2**: FW rep 1 cDNA; **Lane 3**: FW rep 1 -RT; **Lane 4**: PBS rep 1 cDNA; **Lane 5**: PBS rep 1 -RT; **Lane 6**: BHIS rep 1 cDNA; **Lane 7**: BHIS rep 1 -RT; **Lane 8**: FW rep 2 cDNA; **Lane 9**: FW rep 2 -RT; **Lane 10**: PBS rep 2 cDNA; **Lane 11**: PBS rep 2 -RT. **(B) Lane 1**: 100 bp molecular weight marker. **Lane 2**: BHIS rep 2 cDNA; **Lane 3**: BHIS rep 2 -RT; **Lane 4**: FW rep 3 cDNA; **Lane 5**: FW rep 3 -RT; **Lane 6**: PBS rep 3 cDNA; **Lane 7**: PBS rep 3 -RT; **Lane 8**: BHIS rep 3 cDNA; **Lane 9**: BHIS rep 3 -RT; **Lane 10**: gDNA; **Lane 11**: No template control.

### 3.3.4 Primer validation

Primers were designed using Primer 3 Plus software (Table 2.1 and 2.3) as described in Chapter 2 and validated by PCR using genomic DNA from *C. difficile* 630 (Figure 3.5 a and b). Validation of the primer sets shows that the primers are able to bind to the template DNA and produce an amplicon of the expected size when visualised by gel electrophoresis, prior to their use in RT-qPCR.

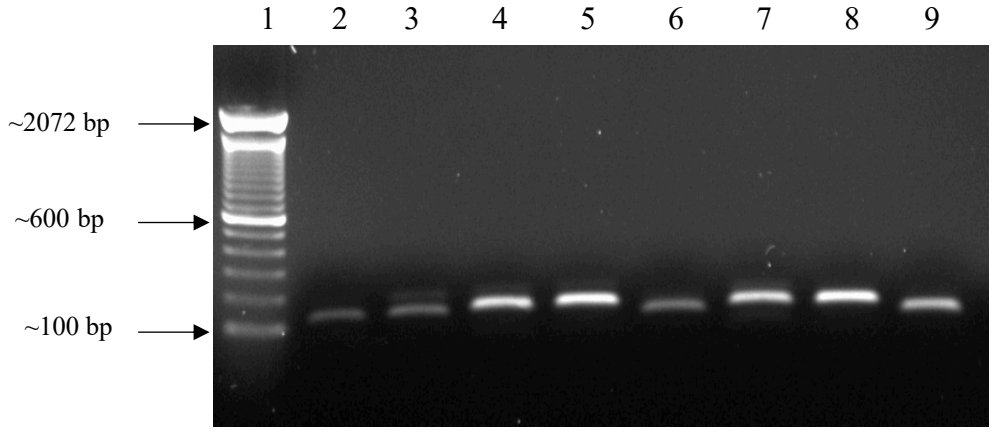



**Figure 3.5a:** Gel electrophoresis image of a selection of primers validated by PCR with *Clostridioides difficile* 630 genomic DNA. **Lane 1**: 100 bp molecular weight marker. **Lane 2**: scdf2295.1; **Lane 3**: scdf3815.1; **Lane 4**: scdf1763.1; **Lane 5**: scdf1153.1; **Lane 6**: *groES*; **Lane 7**: *groEL*; **Lane 8**: *dnaK*; **Lane 9**: *dnaJ*.

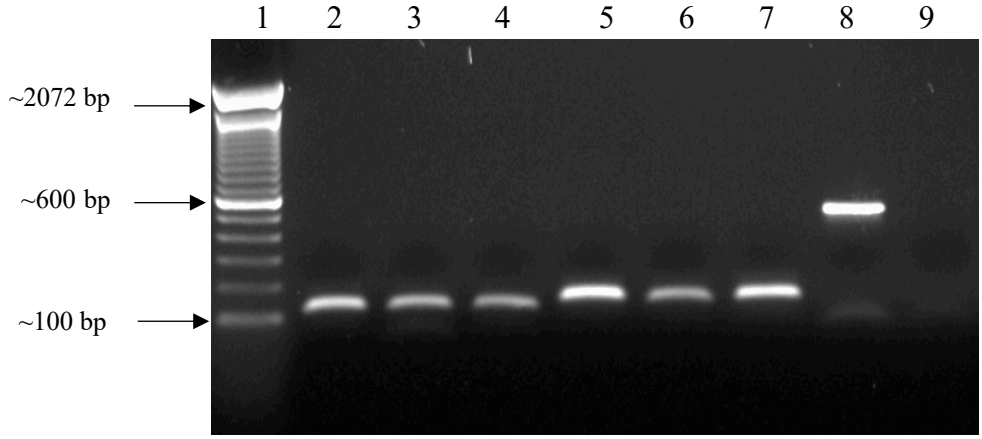



**Figure 3.5b:** Gel electrophoresis image of a selection of primers validated by PCR with *Clostridioides difficile* 630 genomic DNA. **Lane 1**: 100 bp molecular weight marker. **Lane 2**: *fliC*; **Lane 3**: *fliD*; **Lane 4**: *tig*; **Lane 5**: *grpE*; **Lane 6**: *tcdA*; **Lane 7**: *tcdB*; **Lane 8**: *tpi* positive control; **Lane 9**: *tpi* no template control.

### 3.3.5 Reference gene selection and optimisation

In order to ensure the accuracy of subsequent RT-qPCR analysis it was determined that a number of reference genes would be tested and the most efficient three would be chosen for all further qPCR analysis. Metcalf *et al.* (2010) previously investigated the use of eight reference genes in qPCR experiments within a number of *Clostridioides difficile* ribotypes. I decided to assess seven (*adK*, *rpsJ*, *rho*, *tpi*, *gyrA*, *rpoA* and *gluD*) of these reference genes for stability within my *C. difficile* cDNA samples. Validation of the primer sets was initially carried out by end-point PCR (Figure 3.6) and from this, five (*adK*, *rpsJ*, *tpi*, *gyrA*, and *gluD*) were selected for further analysis. Standard curves were carried out and from the data obtained, in addition to the use of BestKeeper software (Table 3.2), it was determined that the three most suitable reference genes to use in all subsequent analysis would be *adK*, *gyrA* and *rpsJ*.

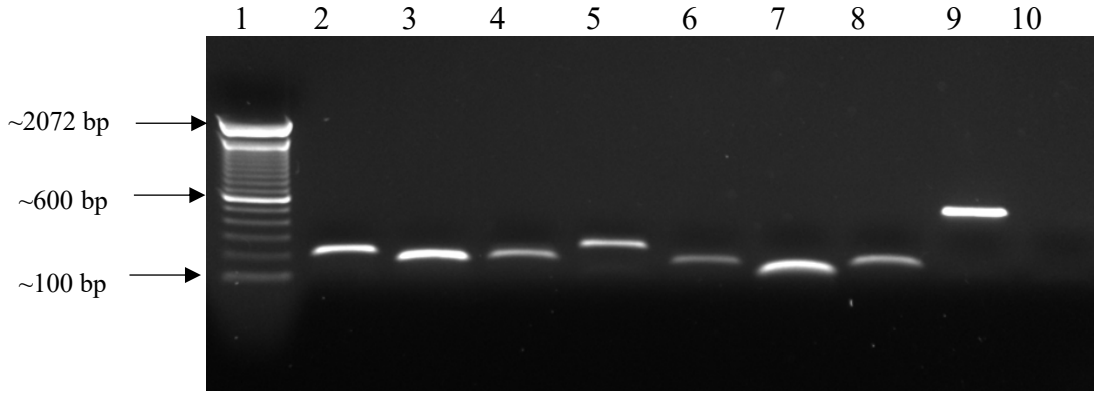



**Figure 3.6:** Gel electrophoresis image of reference gene primer validation by PCR with *Clostridioides difficile* 630 genomic DNA. **Lane 1**: 100 bp molecular weight marker. **Lane 2**: *adK*; **Lane 3**: *rpsJ*; **Lane 4**: *rho*; **Lane 5**: *tpi*; **Lane 6**: *gyrA*; **Lane 7**: *rpoA*; **Lane 8**: *gluD*; **Lane 9**: *tpi* positive control; **Lane 10**: *tpi* no template control.

**Table 3.2:** Analysis of reference genes using BestKeeper software. A standard deviation of < 1 suggests the reference gene is stable enough to be used in downstream RT-qPCR analysis.

| | **Error** | **Efficiency** | **Slope** | **Cq (GEOMEAN)** | **SD (±)** |
|---|---|---|---|---|---|
| ***tpi*** | 0.0487 | 1.905 | -3.572 | 25.44 | 1.13 |
| ***adK*** | 0.0437 | 1.941 | -3.471 | 24.46 | 0.60 |
| ***gyrA*** | 0.0542 | 1.998 | -3.327 | 21.81 | 0.29 |
| ***rpsJ*** | 0.0250 | 1.903 | -3.580 | 19.85 | 0.47 |
| ***gluD*** | 0.0582 | 2.028 | -3.256 | 19.39 | 1.16 |

### 3.3.6 RT-qPCR optimisation

A robust set of guidelines have been developed by Bustin *et al*. (2009) these are known as the MIQE guidelines (Minimum Information for Publication of Quantitative Real-Time PCR Experiments). The purpose of this is to ensure the quality and integrity of published RT-qPCR data in the literature. In order to comply with these guidelines researchers must publish specific information in relation to their experiments such as nucleic acid extraction method and quality, detailed RT-qPCR protocol and data analysis method. I can confirm that these guidelines have been adhered to as far as possible throughout the course of this study.

To evaluate the performance of a primer set and thereby ensure the quality of the RT-qPCR data produced was of a high enough standard for publication the standard curves produced for each gene were assessed using the following criteria: Error: < 0.2; Efficiency: 1.8 to 2.1; Slope: -3.6 to -3.1. Only once these criteria had been met were the standard curves used in further RT-qPCR analysis. Table 3.3 shows the standard curve data produced for genes studied within this chapter and shows that all values fall within the specified parameters. Figures 3.7a, b and c show representative images of the output obtained for a single gene from a typical RT-qPCR experiment.

**Table 3.3:** Standard curve data for all genes studied by RT-qPCR in this chapter.

| | **Error** | **Efficiency** | **Slope** | **Y-intercept** |
|---|---|---|---|---|
| ***adK*** | 0.0437 | 1.941 | -3.471 | 23.35 |
| ***gyrA*** | 0.0542 | 1.998 | -3.327 | 20.08 |
| ***rpsJ*** | 0.0250 | 1.903 | -3.580 | 16.41 |
| ***groEL*** | 0.1520 | 2.073 | -3.159 | 18.15 |
| ***groES*** | 0.0367 | 2.005 | -3.309 | 15.80 |
| ***fliC*** | 0.1270 | 2.079 | -3.146 | 18.73 |
| ***fliD*** | 0.0753 | 1.896 | -3.599 | 21.77 |
| ***tcdA*** | 0.0373 | 1.984 | -3.360 | 23.04 |
| ***spoVB*** | 0.0406 | 2.006 | -3.308 | 28.62 |
| ***spoIIIAC*** | 0.0458 | 1.903 | -3.580 | 25.50 |
| ***cspC*** | 0.0415 | 2.037 | -3.237 | 24.43 |

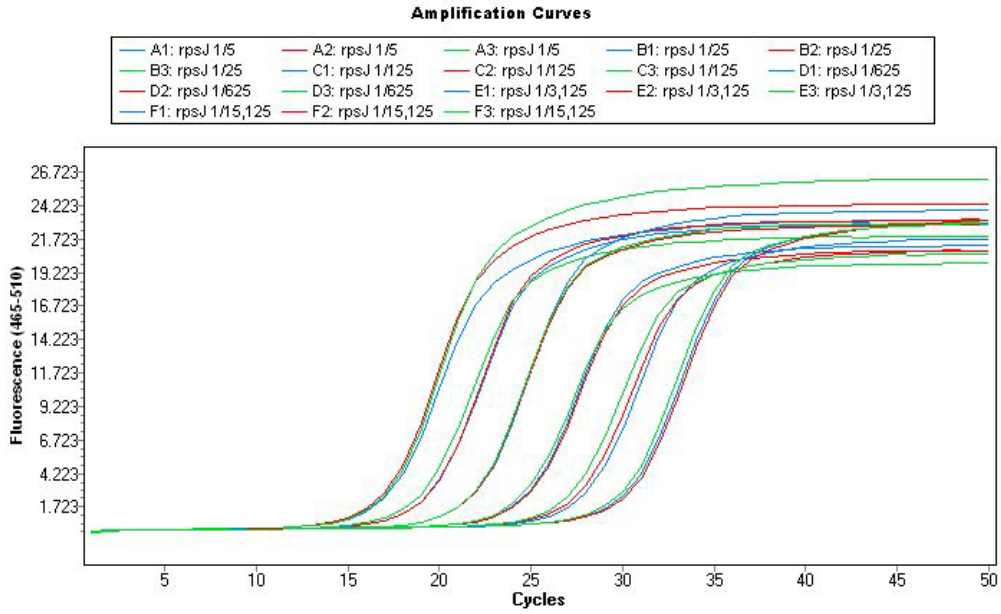
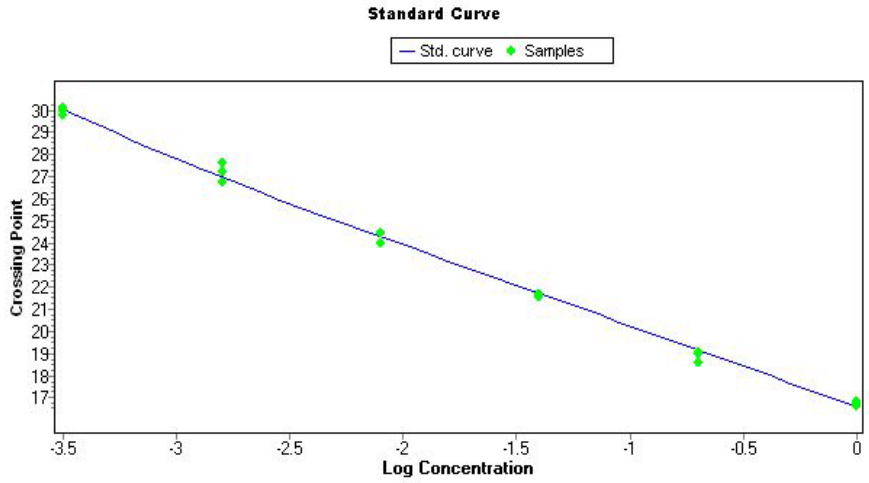



**Figure 3.7a:** RT-qPCR amplification curve for the reference gene *rpsJ* using 5-fold dilutions of pooled cDNA as template.




Error: 0.0166

Efficiency: 1.875

Slope: -3.663

**Figure 3.7b:** RT-qPCR standard curve for reference gene *rpsJ* using 5-fold dilutions of pooled cDNA template.

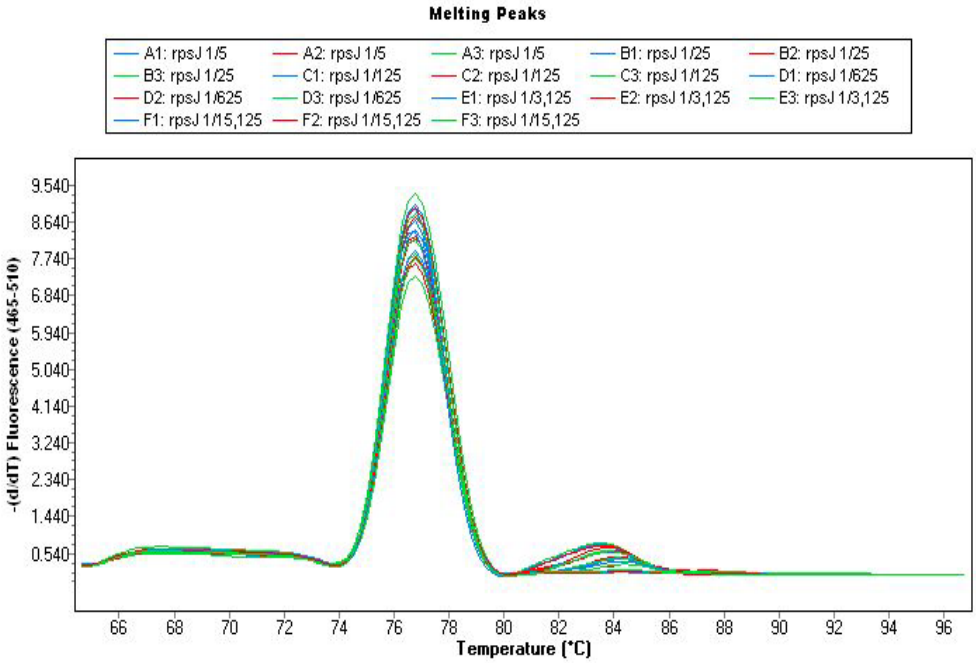



**Figure 3.7c:** RT-qPCR melting curve for reference gene *rpsJ* amplified from 5-fold dilutions of pooled cDNA template.

### 3.3.7 The presence of FW during growth results in modulation of the transcriptome

RNA sequencing found 1218 *C. difficile* 630 transcripts differentially expressed (DE) (padj < 0.05, fold–change (FC) > 1.5) from growth of the organism in the presence of faecal water with 642 (53%) being upregulated and 576 (47%) being downregulated. The highest number of DE genes were categorised as “Similar to unknown proteins” (17.16%),“Transport binding proteins and lipoproteins” (14.94%), Metabolism of amino acids and related molecules” (6.81%), “Transposon and IS function” (5.91%), “Specific metabolic pathways” (5.67%), “Regulation of RNA synthesis” (5.67%) and “Sporulation” (5.5%) (Figure 3.8). RT-qPCR analysis of a selection of genes from the

RNAseq data validated these results. Good correlation ($R^2 = 0.95$) is seen between the two data sets (Figure 3.9) when applied to a number of motility and sporulation genes.

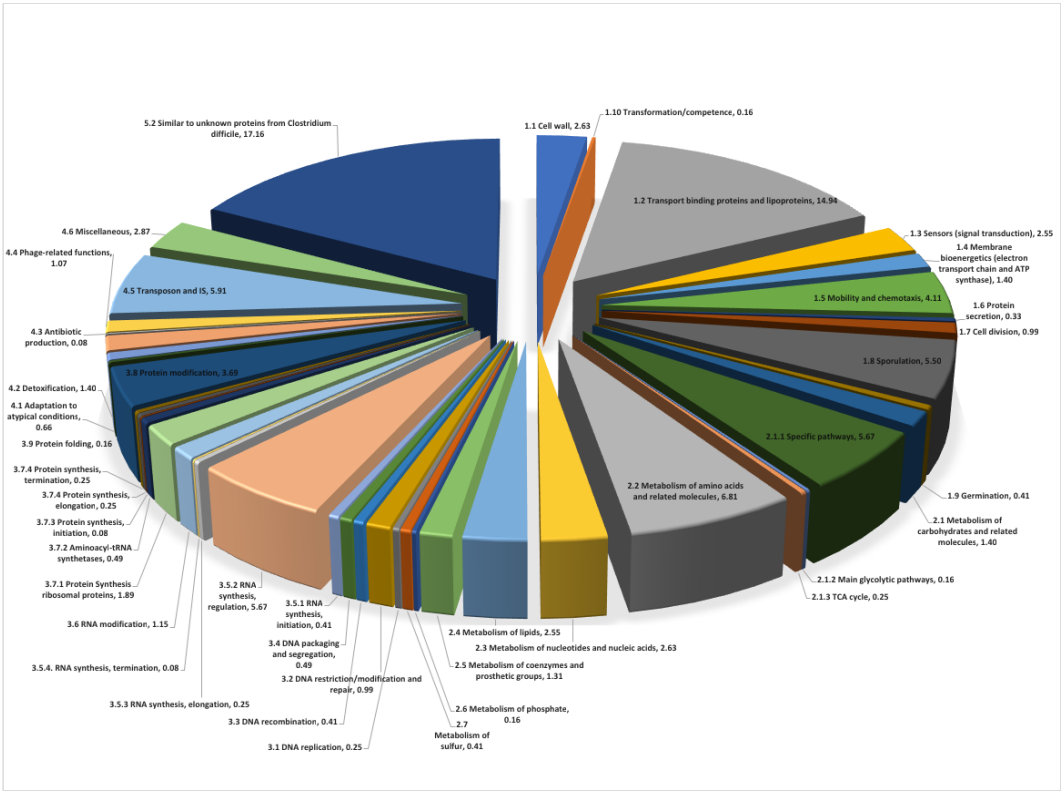



**Figure 3.8:** Functional categorisation of differentially expressed transcripts in the *Clostridioides difficile* 630 faecal water transcriptome. P-value < 0.05, fold change >1.5.

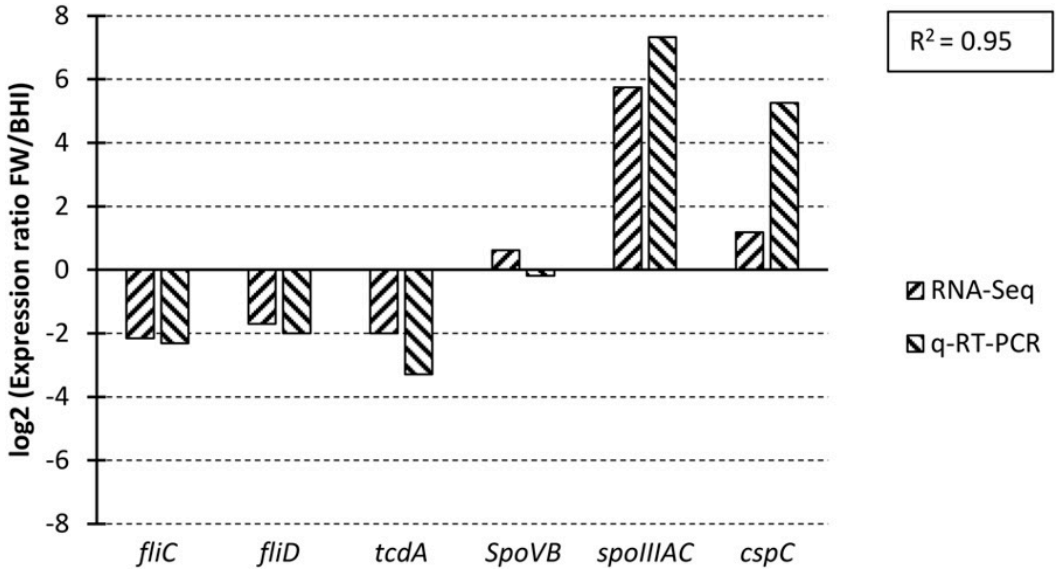



**Figure 3.9:** Comparison of RT-qPCR and RNAseq data for selected *Clostridioides difficile* 630 genes. For each individual gene, expressional changes determined by RNAseq (up-hatched columns) and by RT-qPCR (down-hatched columns) are shown relative to the BHIS control, and show good correlation between the two datasets ($R^2$=0.95). *rpsJ*, *gyrA* and *adk* were used as reference genes.

It was found that a slightly larger proportion of the 1218 DE genes showed increased expression in FW media apart from those genes in the categories of 'motility', 'phage-related functions', 'metabolism of co-factors', 'transport binding' and 'genes associated with specific pathways' (Figure 3.10). The categories where the most significant changes in the transcriptome were seen included 'sporulation', 'motility' and 'similar to unknown proteins from *Clostridioides difficile*'.

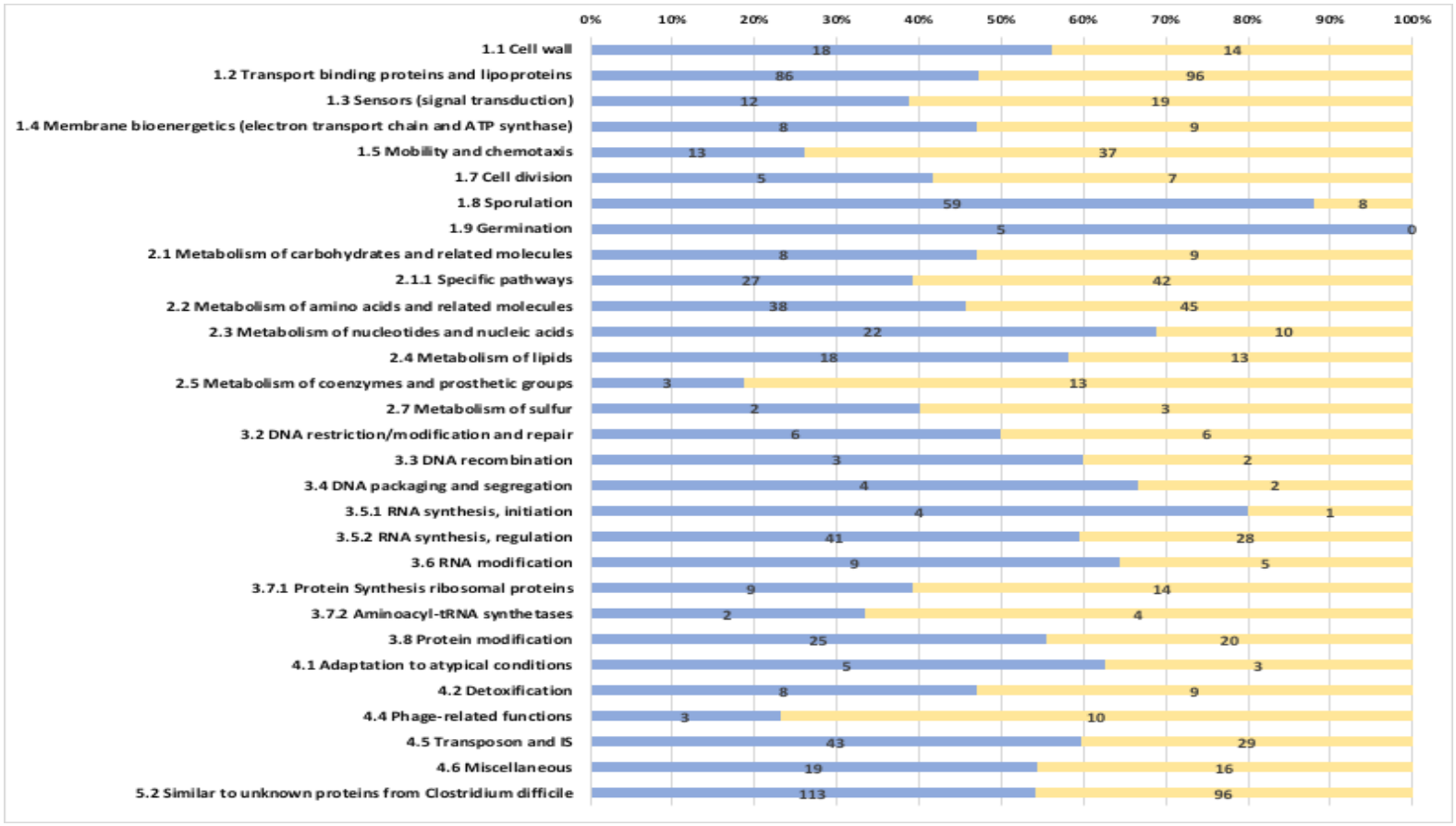



**Figure 3.10:** Proportion of differentially expressed genes in the *Clostridioides difficile* 630 faecal water transcriptome. Blue – up regulated; Orange – down regulated, $p < 0.05$, fold-change > 1.5. Analysis limited to functional categories in which > 5 genes were differentially regulated.

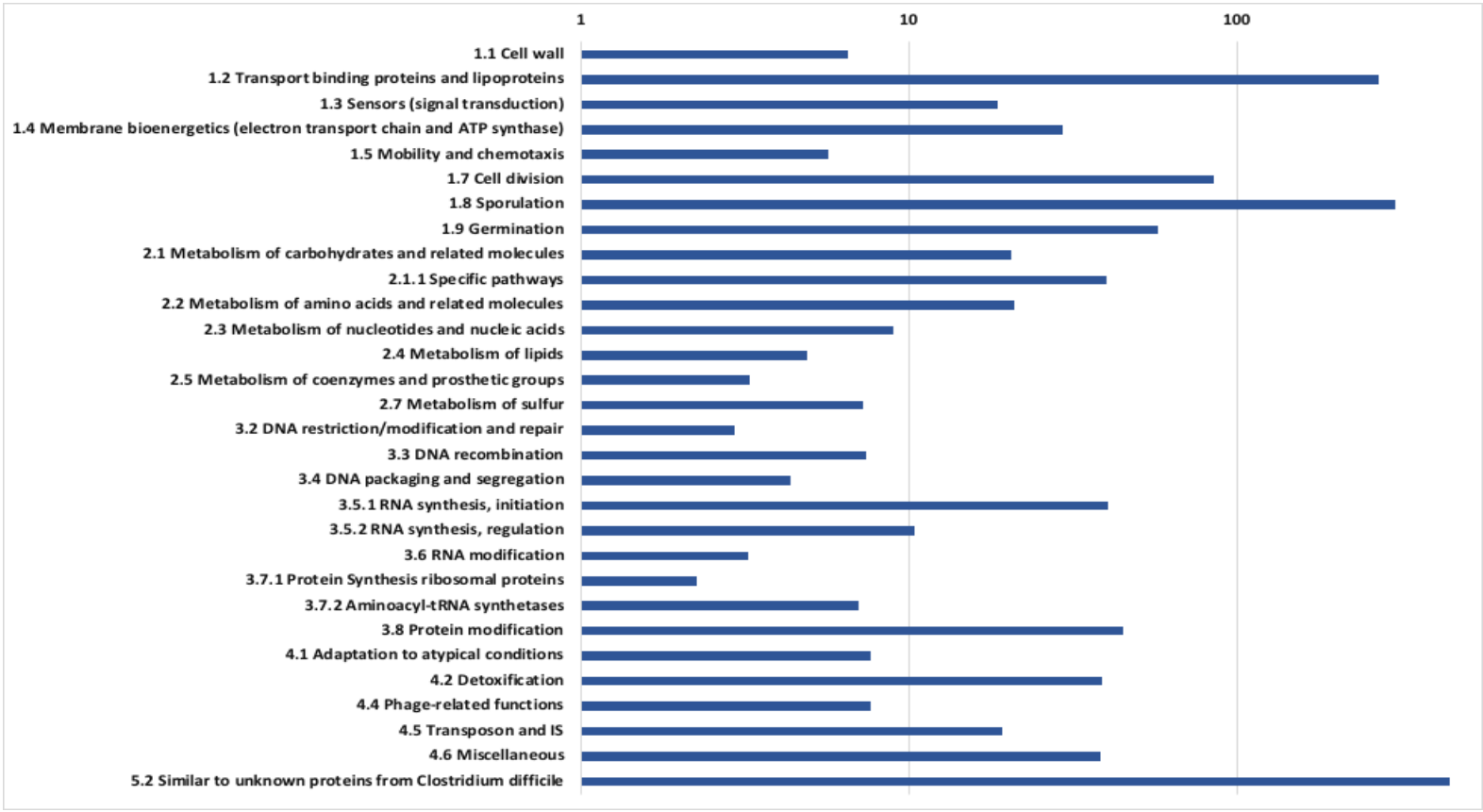



**Figure 3.11:** Magnitude of maximum fold-change in expression of genes within selected functional categories. Transport, sporulation and conserved hypothetical protein encoding genes exhibited noticeably larger changes in expression in the faecal water exposed *Clostridioides difficile* 630 transcriptome. Functional categories with less than 5 DE genes were not included in the analysis.

More than 100-fold increase in expression was observed for nine sporulation-associated genes. The largest increase in expression of 445-fold was seen in CD1065, which encodes a 146 amino acid 'conserved hypothetical protein'. Three genes encoding an ATP binding cassette (ABC) transporter operon (CD0873-0875) showed the greatest decreases in expression (> 100-fold). The categories with the largest fold-changes in expression were transport binding proteins and lipoproteins (up to 270-fold), sporulation (up to 300-fold), and genes encoding hypothetical proteins (up to 445-fold). Fold-changes of < 20 were observed for genes associated with RNA metabolism, specific metabolic pathways, protein modification, adaptation to atypical conditions and miscellaneous categorised genes. Amongst the genes that showed the least change were those involved with metabolism of sulphur, lipids, and coenzymes (Figure 3.11), suggesting that these significant metabolic pathways were reasonably unaffected by the presence of FW.

### 3.3.8 Sporulation

In an attempt to further support the sporulation changes observed from the RNAseq and RT-qPCR data, spores where stained and counted in samples taken each hour throughout the growth cycle (10 fields of view per slide). Unfortunately, this did not correlate well with the transcriptomic data as no significant changes were seen (Figure 3.12). It was noted that the error bars on Figure 3.12 were unusually large, this may be due to insufficient sample volume or inadequate mixing of sample prior to pipetting onto the glass slide. To avoid this in future I would increase the volume of sample and pipette to mix thoroughly to minimise clumping of spores before adding to slide. There may be a number of reasons why the physiological status of the cell does not reflect

the sporulation transcriptome changes, which will be discussed further in the next section.

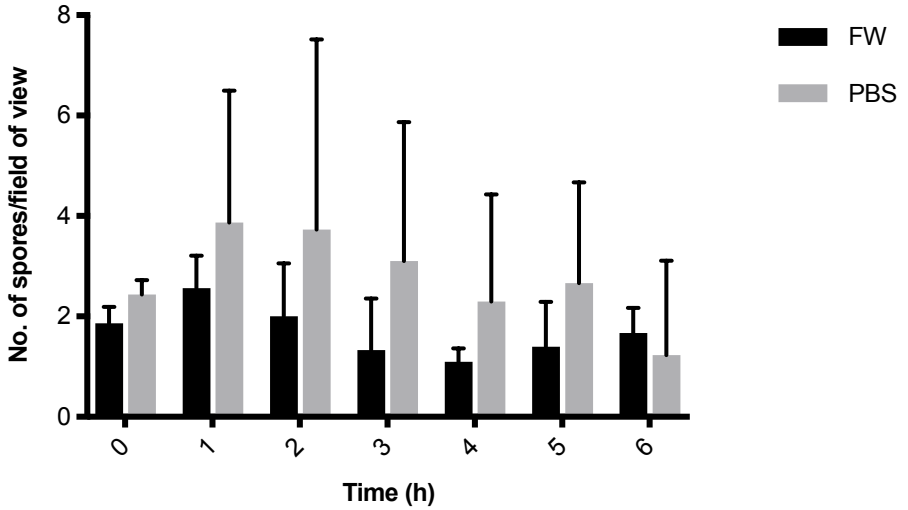



**Figure 3.12:** No. of spores in *Clostridioides difficile* 630 grown in BHIS containing 50% faecal water and in BHIS containing 50% PBS. Samples collected from growth curves shown in Figure 3.1 were stained using malachite green and 10 fields of view were analysed per sample. The data presented are the means of 3 biologically independent experiments and error bars represent standard deviation of the mean. Anova showed no significant change.

### 3.3.9 Cell length

It was found that bacterial cell length increased by almost 70 % at 6 h in FW media, with an average cell length of 4.3 µm versus 3.3 µm in the control media ($p < 0.05$) (Figure 3.13), this was despite the growth rate remaining unchanged between media types as seen in Figure 3.1.

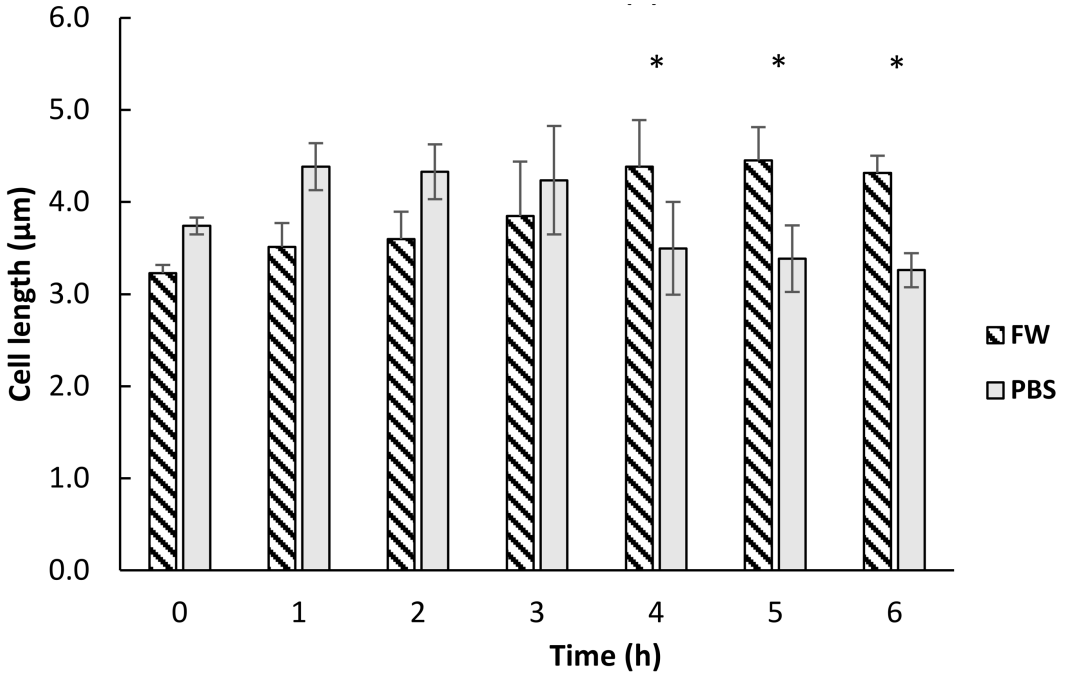



**Figure 3.13:** Cell length of *C. difficile* 630 grown in BHIS/faecal water (50/50v) and BHIS/PBS (50/50v). Samples from growth curves shown in Figure 3.1 were Gram's stained by standard methods and a total of 100 cells were measured per sample. The data presented are the means of 3 biologically independent experiments and error bars represent standard deviation of the mean. *$p < 0.05$, Anova, Post hoc: Dunnett t (2-sided).

## 3.4 DISCUSSION

### 3.4.1 RNA extraction and QC

Reliable gene expression studies require high quality RNA but due to the Gram-positive nature of *C. difficile* it can be difficult to extract RNA that is of a quantity and standard suitable for downstream applications such as RNA sequencing, cDNA synthesis, RT-qPCR and Northern blotting. The thick cell wall of Gram-positive organisms can make it difficult for many chemical lysis methods to penetrate, resulting in low yield (Nuyts *et al.*, 2001). Over-vigorous mechanical lysis on the other hand can lead to shearing and loss of genetic material resulting in a low RNA integrity number (RIN) and/or poor purity values due to residual contaminants (Metcalf and Weese, 2012). We were able to extract not only high yields but also good quality RNA by optimising a commercially available RNA extraction kit (RNeasy® Mini kit, Qiagen, UK); this has resulted in the establishment of an effective and straightforward approach for the isolation of high quality RNA from *C. difficile* 630.

### 3.4.2 RT-qPCR and reference gene optimisation

It is widely accepted that RT-qPCR is one the most reliable and efficient methods for quantifying the level of nucleic acid expression within a sample (Bustin, 2000). The MIQE guidelines were developed in 2009 by Bustin *et al*. as a means of standardising RT-qPCR experiments and the accurate publication of the results obtained. In addition to high quality RNA as previously discussed the use of suitable stable reference genes forms a main part of the guidelines due to importance of precise normalisation of relative gene expression levels which allows consistent interpretation of RT-qPCR data. One particularly well known use of inaccurate RT-qPCR data involves the

publication of a study purporting to show a link between autism and the measles, mumps and rubella (MMR) vaccine (Wakefield *et al*., 1998), the article was ultimately retracted due to the discovery of inaccurate qPCR data, but not before impacting greatly on worldwide uptake of the vaccine, putting many hundreds of thousands of vulnerable children at risk. The paper stated that the measles virus (MeV) had been detected in the intestine of children that had been diagnosed with autism, however no RT-qPCR data was included to support this claim or allow these results to be reproduced independently. In addition to this, there was obvious signs of contamination within the negative controls which was ignored. Furthermore, there was no mention of the quality of the RNA being analysed; this would be essential information in order to assess the legitimacy of the results. Only a single reference gene was used and no standard curve data was made available. In all, the resultant qPCR data analysis was deemed wholly unreliable (Bustin, 2013).

The work carried out by Metcalfe *et al*. (2010) into the use of reference genes within *C. difficle* qPCR allowed testing of a number of reference genes for suitability with my samples. This has led to the development of a trio of reference genes (*adK*, *rpsJ* and *gyrA*) that can be used in future RT-qPCR analysis of *C. difficile* 630 experiments whilst safe in the knowledge that accurate normalisation is being achieved.

### 3.4.3 Sporulation

It is known that in *Bacillus subtilis* sporulation is a tightly controlled process involving sporulation–specific sigma factors (σH (early), σF, σE, σG, and σK (late)) which are

activated at highly specific stages of the sporulation process (Underwood *et al*., 2009; Al-Hinai *et al*., 2014; Abt *et al*., 2016). Browne *et al*. (2016) have recently shown extensive sporulation capability within the human gut microbiota, we therefore hypothesised that FW would induce sporulation in *C. difficile*. We found that sporulation associated genes were amongst the most differentially expressed in the FW grown samples, with up to 300-fold increases. From 67 sporulation associated genes that were differentially expressed in the sequencing data, there were just eight down-regulated in the FW grown cells, these included CD2273 encoding a 'putative sporulation integral membrane protein' which may be under the control of σE, and CD3669 encoding a 'putative exported protein' which is part of the mature spore proteome.

Nineteen of the 59 genes remaining were upregulated by more than 50-fold – this strongly suggests that FW is an extremely powerful inducer of sporulation in *C. difficile* 630. Sporulation in *C. difficile* has been extensively studied which has allowed the genes under the control of the master regulator Spo0A (CD1214 - increased by 2.5-fold) to be further characterised, thereby facilitating the identification of possible connections between sporulation and other phenotypes (Lawley *et al*., 2009; Pettit *et al*., 2014; Saujet *et al*., 2014; Browne *et al*., 2016). A sigma factor cascade is initiated by phosphorylation of the Spo0A protein, this impacts on the expression of the sigma factors σE (CD2643, 23-fold up), σG (CD2642, 40-fold up), σF (CD0772, 9-fold up) and σK (CD1498, 1.8-fold down) within both the mother cell and forespore, and controls expression of early (stage II and III) and late (stage V and VI) sporulation genes (Pettit *et al*., 2014). Asymmetric division of the cell creates two unevenly sized

sections during sporulation due to the development of a septum – the smaller section (the forespore) becomes the spore and the larger section prepares the forespore for dormancy (Abt *et al*., 2016).

Thereby, our data would suggest that the FW grown *C. difficile* 630 cells, at the point in which they were collected were physiologically at stage III of sporulation, this is the stage at which engulfment of the forespore has occurred, but before cortex formation. This is supported by previous studies into the identification and analysis of sporulation-related genes (Pettit *et al*., 2014; Browne *et al*., 2016). This correlates with the expression of the *spoIIIAA–spoIIAH* operon (all > 50-fold up), in addition to *spoIIIJ* (oxaA1, 1.6-fold up), *spoIIID* (56-fold up) and *sigG* (40-fold up) which were all significantly up-regulated in the FW grown samples. The sigma factor with the highest increase in expression was σE at 23-fold up in the FW samples. As the protein that σE produces is known to exert its effects on a number of genes present within the clostridial sporulation cascade, it was of note that there was an increase in expression of σE controlled genes including *spoIIID*, *spoIVA* (57-fold increase), *spoVD* (29-fold increase), *cspBA* (22-fold increase) and *cspC* (2.2-fold increase) within the RNAseq dataset (Pettit *et al*., 2014; Abt *et al*., 2016). In addition to this, there was an increase in expression of genes known to encode spore coat proteins including CotE (CD1433) which was 29-fold up in our data, this is significant as the chitinase and peroxiredoxin properties of CotE are known to facilitate pathogenesis by instigating the breakdown of gut mucus during infection (Permpoonpattana *et al*., 2013; Paredes-Sabja *et al*., 2014).

Edwards *et al*. (2014) have indicated that a reduction in the expression of oppABC (CD0853-855) which encodes an oligopeptide transporter, contributes to earlier expression of sporulation-associated genes, this is in keeping with our data where oppABC expression was 50% lower in FW and therefore consistent with other indications of FW induced changes in the *C. difficile* transcriptome within our data. It can therefore be suggested that the data collected shows that cells grown in FW media are sporulating more rapidly than those grown in BHIS media. These observations have strong implications for pathogenesis of *C. difficile* especially as spores are the infectious, transmissible form of the organism, and consistent with the observations of other studies on the extent of sporulation of microbiota within the gut (Browne *et al*., 2016).

### 3.4.4 Transport systems

Transporter gene expression underwent considerable modulation within the FW grown cells, this is in contrast to previous work which showed that phosphotransferase (PTS) sugar transport systems were mostly unaffected by heat-stress (Ternan *et al*., 2014). The PTS consists of two main components - enzyme I (EI), and the histidine phosphor-carrier protein (HPr) and is the key bacterial carbohydrate assimilation system for hexoses, hexitols and disaccharides in the cell membrane. Our dataset shows an increase of 1.58-fold in the expression of the gene encoding the EI element (CD2755) which is a critical component of all phosphotransferase systems in the cell. However, the gene HPr kinase/phosphorylase (CD3409) was 1.8-fold down, this gene is known to phosphorylate the cytoplasmic phosphocarrier protein Hpr at Ser42 and leads to initiation of the LacI family carbon catabolite repressor, ccpA (Martin-Verstraete,

2016). In addition to these findings, the genes bglF (CD0388) which encodes the IIABC component of the PTS system for uptake of beta-glucosides was up-regulated as well as the downstream gene bglA (CD0389) encoding 6-phospho-beta-glucosidase, this may be due to the increased presence of glucoside substrates in the FW media compared to the BHIS control media (Williamson and Clifford, 2017). Furthermore, the sortbitol specific component srlEa (CD0765), the fructose specific IIABC component, fruABC (CD2269), and the genes encoding the IIA (CD2512) and IIB (CD2510) components of the glucose PTS transport system were both found to be increased in the FW media. In addition to this, the IIC (CD3277) and IID (CD3276) components of the mannose/fructose/sorbose transport system were up by 4 and 6.6-fold, respectively. However, the data shows that the genes within the *xyl* and *xyn* operons (CD3064-CD3070) which are PTS system components and involved in the uptake of xylosides were decreased in the FW media, whereas *xylR* (CD3066) a transcriptional regulator and xylose repressor which works to lower expression of genes for uptake and metabolism of xylose showed increased expression in FW. It could be suggested that these observations are the organisms attempt to adapt and respond to the increase in carbon sources and other diet derived metabolites that are present in the FW media. The PTS is also a signalling mechanism which researchers have been able to connect to chemotaxis and regulatory capabilities related to C, N and P metabolism as well as the virulence of *C. difficile* (Antunes *et al*., 2011; Deutscher *et al*., 2014). The intricate relationship between a number of cellular processes such as carbon catabolite repression, quorum sensing, sugar transport and amino acid metabolism has been shown to regulate toxin production, we know for example that butyrate stimulates toxin synthesis however in the FW grown samples we found a reduction in the expression of 12 genes linked to carbohydrate fermentation

to butyrate (Martin-Verstraete *et al*., 2016). This is consistent with the decreased expression of *tcdA* (CD0663) by 4-fold in FW media and suggests a lowered affinity for butyrate production within the gut environment.

We also noted that a number of genes encoding ABC transporters were DE in FW media, this included many involved in the transport of vitamins, amino acids, sugar phosphates, oligopeptides as well as transporters linked to multidrug efflux mechanisms (Suárez *et al*., 2013). The gene with the lowest expression at 270-fold down in FW was an ABC transporter 'substrate-binding lipoprotein' (CD0873), which acts as an adhesin that allows *C. difficile* to bind to Caco-2 cells (Kovacs-Simon *et al*., 2014). Therefore, we could hypothesis that there is reduced binding to epithelial cells at the point of cell collection, in addition to the reduced motility and increased sporulation, and thereby suggest that this physiological condition could enable removal of the organism from the infected host. The data also showed increased expression of the lantibiotic / multidrug ABC transporters CD0161, CD1349, CD1350 CD2210, CD2211. *C. difficile* as well as other gut pathogens encounter many innate host defences such as cationic antimicrobial peptides (CAMPs) produced by both host and native microbiota (Deutscher *et al*., 2014; Suárez *et al*., 2013). Previous research has shown that the proteins cprA (ATP–binding protein) and cprB (permease) encoded by CD1349 and CD1350 respectively are involved in resistance to CAMPs (McBride and Sonenshein, 2011). This is consistent with our initial hypothesis that increased levels of antimicrobial peptides present in FW lead to increased expression of this mechanism. We can also deduce that the changes we have observed in gene expression

are more than likely as a result of the presence of FW as there was very little change in expression of the classical stress-response genes.

### 3.4.5 Flagella, motility and chemotaxis

Flagella are molecular apparatus with the ability to self-assemble, their synthesis is a tightly controlled process with a highly ordered control of gene expression ensuring the production of later-stage parts is inhibited until the assembly of earlier parts is finished. Flagellin (FliC) for example is only expressed after the basal body and motor machinery are completed. In combination with type-IV pili they are heavily involved in the pathogenesis of a number of enteropathogens (Pallen and Matzke, 2006; Deligianni *et al*., 2010; Dingle *et al*., 2011; Evans *et al*., 2014; Baker, 2016). Our data showed that genes located in the F3 loci (CD0245-CD0271), including those encoding constituents of the basal body, motor, hook and rod where 1.5 to 2-fold down-regulated in the FW media. Similarly, the genes in the F1 locus (CD0226-CD0240) were down-regulated by 2 to 5.7-fold – this could be due to a reduced requirement for these proteins until following construction of the basal body. Also, FliC (CD0239), FlgN (CD0230) and genes in the interflagellar F2 locus (CD0240-CD0244) were also down by between 3 and 5.7-fold. Thus, flagellar operon gene expression and therefore motility of *C. difficile*, decreases in FW alongside increased expression of sporulation genes. Dingle *et al*. (2011) found that a *C. difficile fliC* mutant strain showed improved adherence to intestinal-derived Caco-2 cells suggesting that flagella are either not needed for virulence or that reduction of motility is a tool utilised to increase pathogenesis of the organism.

Genes encoding a type IV pilin, a type–II secretion system protein, and a pilus assembly ATPase (CD3294-6) were up-regulated by 2-3-fold, this may indicate that pilus-driven motility could be of more significance in a FW environment and therefore at certain points of the infection cycle. It is known that flagella are involved in the development of intestinal lesions through the host inflammatory responses, as TLR5 is recognised by FliC which leads to NF-κB and the MAPK signalling pathway activation and production of proinflammatory cytokines (Batah and Kansau, 2016). As these types of receptor are not present in our experimental system the reduction in flagellar expression may suggest a complex relationship between a putative motility phenotype and an adhesion, or possibly, sporulation phenotype. It could be suggested that in the intestinal environment and presence of semi-solid material, as a form of energy conservation the cell may be propelled by a less resource demanding structure such as type IV pili.

### 3.4.6 Genes encoding conserved hypothetical proteins

From the data obtained the greatest number of DE genes, 209 in total, were in the category of 'Similar to unknown proteins', from this category 19 were DE by more than 20-fold. Unfortunately none of the protein products of these genes had predicted signal peptides and were all predicted by SecretomeP to be non-classically secreted, apart from CD1726 and CD3522 (Petersen *et al*., 2011; Bendtsen *et al*., 2005). From PsortB nine gene products had predicted locations in the cytoplasmic membrane with the remaining proteins possessing no conserved domains that might suggest their possible function (Yu *et al*., 2010).

However, extensive literature and database searches has led to the identification of many to an involvement in sporulation pathways. The expression of CD2344 was the most reduced in the FW media of all the conserved hypothetical protein-encoding genes. The literature states that this gene is a putative succinate transporter involved gut colonisation in *C. difficile* (Ferreyra *et al*., 2014). A number of other genes in the succinate to butanoate fermentation pathway, which are downstream of CD2344 (in the same operon) such as *cat1*, *sucD*, *abfD*, *cat2*, and *4hbd* were also down regulated in the FW media by approx. 4-fold (Karp *et al*., 2016). Many genes thought to be regulated by sporulation-associated sigma factors including CD3850 and CD1065 (σK), CD2808 and CD2375 (σG), CD1063A-C, CD2150A and CD3522 (σE) were also found to be DE in FW media (Pettit *et al*., 2014; Saujet *et al*., 2014, Pishdadian *et al*., 2015). It has also been observed by Dembek (2014) that a large number of *C. difficile* spore transcripts encode proteins of unknown function, they thereby suggested that this was due to the difference between the transcriptional status of vegetative cells and that of spores.

We can conclude that although FW did not affect growth rate, an increase in cell length was observed which can be assumed to be a prelude to sporulation. We have shown an acceleration in the sporulation cascade with large increases in the expression of sporulation associated genes, and simultaneous decreases in motility and toxin gene expression – a possible reflection of the relationship between FW and various regulatory systems, and transcription factors, many of which also showed increased expression in FW media. It can be assumed that *C. difficile* adapts well to a faecal environment due to the lack of classical stress response observed. This FW model

signifies a new method of investigation into the adaptability and transcriptomic response of *C. difficile* 630 to the gut setting.

# CHAPTER 4: TRANSCRIPTOMIC ANALYSIS OF *CLOSTRIDIOIDES DIFFICILE* 630 MUTANT STRAINS GROWN IN A FAECAL WATER MEDIUM

## 4.1 INTRODUCTION

Development of the ClosTron system by Heap *et al*. (2007) paved the way for the creation of reproducible mutants in an extensive array of clostridial species. This system utilises mobile group II introns, which make use of the retargeting capabilities of the Ll.ltrB group II intron of *Lactococcus lactis* to generate targeted gene disruptions and the insertion of an erythromycin resistance gene, *ermB*, that allows for positive selection of the resulting mutants (Kuehne *et al*., 2011) (Figure 4.1). Group II introns are retrotransposition elements that are composed of a self-catalytic intron RNA fragment and a multifunctional intron-encoded protein (IEP). Expression of the group II intron within the genome results in the formation of ribonuclear protein (RNP) complex between the Ll.LtrB group II-IEP (LtrA) and the transcribed intron RNA (Heap *et al*., 2007; Heap *et al*., 2010). Subsequent "retrohoming" is initiated through intron mobility and recognition of the target DNA (Zhong *et al*., 2003). Insertion of the RNA into the DNA is followed by reverse transcription to complete formation of the double stranded DNA resulting in disruption of the target gene (Mohr *et al*., 2000; Zhong *et al*., 2003). As explained in Chapter 1, an antibiotic resistance marker (*ermB* – which codes for erythromycin resistance) is incorporated into the design of the group II introns to allow for positive selection of mutants. ErmB is inactivated by the insertion of a group I intron (phage *td*). When the group II intron inserts successfully into the chromosome and is transcribed, the group I intron is able to self-splice out as it is now in the correct orientation to do so, thereby reactivating the erythromycin resistance gene and allowing for positive selection of intron-containing ClosTron disruption mutants on erythromycin-supplemented media (Heap *et al*., 2007) (Figure 4.1).

A number of mutants have now been developed for a wide selection of genes including *dnaK* (Jain, 2010; Jain *et al*., 2017) and *spo0A* (Heap *et al*., 2007) both of which we have utilised within this chapter. Many other disruption mutants have been developed using ClosTron technology across a variety of clostridial species and genes. In *C. difficile* the genes encoding the flagellin subunit, FliC (Twine *et al*., 2009), the sporulation master regulator, Spo0A (Underwood *et al*., 2009), the cell surface protein, Cwp84 (Kirby *et al*., 2009), the cell wall protein, CwpV (Emerson *et al*., 2009), and the putative spore cortex-lytic enzyme, SleC (Burns *et al*., 2010) have all been disrupted to create mutants in order to elucidate gene function. Many other clostridial species have been engineered using the ClosTron system including the genes encoding - the neurotoxin type A, bont/A (Bradshaw *et al*., 2010) and the signalling protein AgrD in *C. botulinum* (Cooksley *et al*., 2010), the peptidoglycan hydrolase, Acp in *C. perfringens* (Camiade *et al*., 2010), and the putative type II restriction endonuclease, Cac8241 in *C. acetobutylicum* (Dong, *et al*., 2010) to name a few.

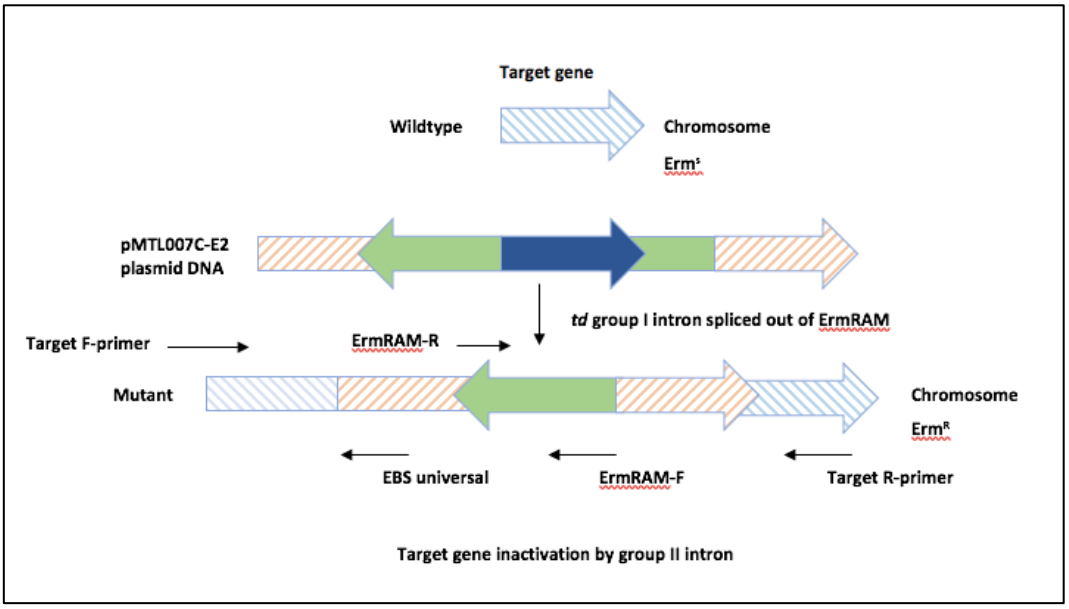



**Figure 4.1:** Schematic representation of the ClosTron mutagenesis system. The pMTL007C-E2 plasmid consists of the Ll.LtrB group II intron (orange hatched arrow) with an internal RAM conferring erythromycin resistance (green arrow), which is disrupted by the *td* group I intron (dark blue arrow). The group II intron is retargeted to the target gene (light blue hatched arrow), insertion of the group II intron into the chromosomal DNA in the correct orientation leads to disruption of this gene. Splicing out of the *td* group I intron from the ErmRAM restores a functional *ermB* gene allowing for the positive selection of mutants. Screening PCR primers are shown as black arrows. ErmS = erythromycin sensitive; ErmR = erythromycin resistant. Adapted from Heap *et al*., 2010.

Development of the *dnaK* disruption mutant by Jain *et al*. (2017) has allowed us to study the effect of the deletion of this heat-shock protein (HSP) on *C. difficile* 630. HSPs are involved in homeostasis, protein folding and refolding and prevention of protein aggregation with the bacterial cell (Young *et al*., 2004; Hartl and Hayer-Hartl 2009; Hartl *et al*., 2011). These functions provide the organism with an innate robustness protecting them from environmental perturbations and allowing them to survive and recover from stressful conditions (Aguilar-Rodriguez *et al*., 2016). The

DnaK protein works to ensure the correct folding of denatured proteins and newly synthesised polypeptide chains as they emerge from the ribosome (Popp *et al*., 2005). DnaK works alongside its co-chaperones DnaJ and GrpE; all three chaperones interact closely, alongside their negative regulator HrcA, to help maintain a steady state within the cell (Schonfeld *et al*., 1995). See Figure 4.2 for a graphical representation of how bacterial chaperones interact to ensure efficiency of protein folding within the cell (Calloni *et al*., 2012).

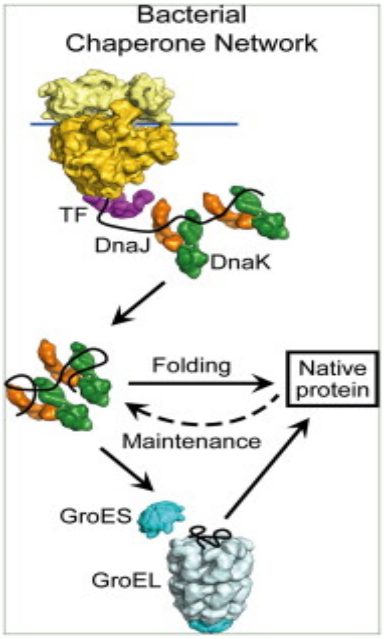



**Figure 4.2:** The bacterial chaperones DnaK, DnaJ, GroES, GroEL alongside Trigger Factor (TF) interact closely with one another to ensure efficient protein folding within the cell (Calloni *et al*., 2012).

A *C. difficile spo0A* disruption mutant was originally reported by Heap *et al*. describing application of the, then novel, ClosTron gene disruption system (Heap *et al*., 2007) allowing further analysis of how sporulation occurs within this organism

(Deakin *et al*., 2012). Sporulation is a multifaceted sequence of events leading to the formation of metabolically inactive spores from vegetative cells (Paredes-Sabja *et al*., 2005). In *C. difficile* and other sporulating bacteria it has been shown that Spo0A is the major transcription factor that functions in the initial stages of sporulation (Chastanet *et al*., 2010) (See Chapter 1, Figure 1.6 for *C. difficile* 630 sporulation cascade proposed by Pettit *et al*., 2014). A non-sporulating phenotype is induced from genetic inactivation of the *spo0A* gene (Heap *et al*., 2007; Underwood *et al*., 2009). In a number of *Clostridium* and *Bacillus* species including *C. difficile* it has been shown that Spo0A binds directly to DNA upstream of a number of early sporulation-associated genes including *spoIIAA*, *spoIIE*, and *spoIIGA* (Rosenbusch *et al*., 2012). Several studies to date have also indicated that this transcriptional regulator is involved in the control of other processes not apparently associated with sporulation such as the expression of the toxins TcdA and TcdB. There is disagreement in the published literature, with *tcdA* and *tcdB* being reported variously as positively regulated in a s*po0A* mutant (Underwood *et al*., 2009), negatively regulated **(**Deakin *et al*., 2012; Mackin *et al*., 2013; Pettit *et al*., 2014) and unaffected by *spo0A* disruption (Rosenbusch *et al*., 2012). A number of contradictions thus currently exist in the literature regarding the role of Spo0A in toxin expression thereby warranting further investigation. Fourteen flagellar proteins including FliC were found to be upregulated at the mRNA level in the s*po0A* mutant by Pettit *et al*. (2014), who also found a significant decrease in glucose metabolism within the *spo0A* mutant and, furthermore, that the entire butyrate production operon was downregulated at both the mRNA and protein levels. Moreover, biofilm formation has been found to be decreased in *C. difficile spo0A* mutant strains grown in BHIS at 37 °C (Dawson *et al*., 2012; Dapa *et al*., 2013).

As discussed in Chapter 3, FW provides a growth environment that can be considered as being similar to the contents of the human gut and therefore allows the identification of potential mechanisms by which *C. difficile* causes infection and disease that would not be observed in normal laboratory media. We have focused in this chapter on analysis of the expression of a number of genes involved in protein folding and cell homeostasis, motility, toxin expression and sporulation, as well as making observations of morphological changes of the mutants in response to FW by measuring cell length.

## 4.2 AIMS AND OBJECTIVES

The aim of this chapter is to investigate the expression of key genes associated with motility, sporulation, toxin production and chaperone activity in four *C. difficile* strains grown in FW media. In particular, we wanted to see what effect, if any, the lack of *dnaK* and *spo0A* expression had on growth and gene expression when grown in a more physiologically-relevant environment such as our FW model.

The main objectives for this chapter are:

- Determine the effect of faecal water on the growth of the *C. difficile* strains 630, *Δerm, dnaK* and *spo0A.*
- Measure effect of faecal water on cell length in the four strains.
- Investigate the effect of faecal water on motility, sporulation, toxin and chaperone gene expression in the mutant strains in comparison to the wildtype.

## 4.3 RESULTS

### 4.3.1 DNA extraction from each of the four *C. difficile* 630 strains

DNA was extracted from each strain, using the Fast DNA Spin Kit (Chapter 2, Section 2.11) in order to confirm the genotype of each one by PCR. Both quality and quantity was measured using Nanodrop. Yields ranged from 194 to 318 ng/ul; 260/280 values ranged from 1.51 to 1.80 and the 260/230 values ranged from 0.60 to 0.75 (Table 4.1). Ideal Nanodrop values are: 260/280 – 1.8 – 2.0 and for 260/230 2.0 – 2.2. Although the values reported here fall outside the ideal ranges no amplification issues were encountered.

**Table 4.1:** Nanodrop values for DNA extracted from each of the four strains from cells collected at late-log phase of growth. Two extractions were performed for each batch of cells collected.

| | ***ng/uL*** | ***260/280*** | ***260/230*** |
|---|---|---|---|
| **630 (1)** | 306.6 | 1.55 | 0.75 |
| **630 (2)** | 223.3 | 1.51 | 0.68 |
| ***Δerm* (1)** | 232.8 | 1.56 | 0.60 |
| ***Δerm* (2)** | 236.1 | 1.61 | 0.63 |
| ***dnaK* (1)** | 259.8 | 1.79 | 0.71 |
| ***dnaK* (2)** | 318.3 | 1.80 | 0.75 |
| ***spo0A* (1)** | 239.4 | 1.68 | 0.64 |
| ***spo0A* (2)** | 194.3 | 1.68 | 0.60 |

### 4.3.2 Confirmation of strain by genotyping PCR

In order to validate the mutant and parental/wildtype strains PCR was performed using the extracted DNA. Genotyping PCR was carried out using the primer sets 4140/4620, 4140/5880, 5880/7020, dnakF/dnakR spo0AF/spo0AR, EBS universal/dnaKR and EBS universal/spo0AR (Figures 4.4 a, b, c and d). For primers used see Chapter 2 Table 2.2. The amplicon size expected in base pairs (bp) for each primer pair in each strain is outlined in Table 4.2. The primer set 4140F/4620R amplifies a 480 bp product specific to the erm1(B) gene and the Erm leader peptide present in all four isolates as expected as this section of DNA is retained in the mutant strains. The primer set 4140F/5880R is designed to amplify a 1740 bp product representing the Erm leader peptide, erm1(B), and ORF3 and ORF298 in *C. difficile* 630 but not in *C. difficile* *Δerm* and its derivatives. Primer set 5880F/7020R (1140 bp amplicon) specific to the ORF298 and erm2(B) was amplified from *C. difficile* 630 but again not in the *C. difficile* strains *Δerm*, *dnaK* or *spo0A* as this section has been lost from these strains. This confirms the 2.4 kb deletion in Tn5398 in *Δerm* (and derivatives) containing most of ORF298 and all of the erm2 (B) gene including some flanking regions as outlined in Figure 4.3 (Hussain *et al*., 2005).

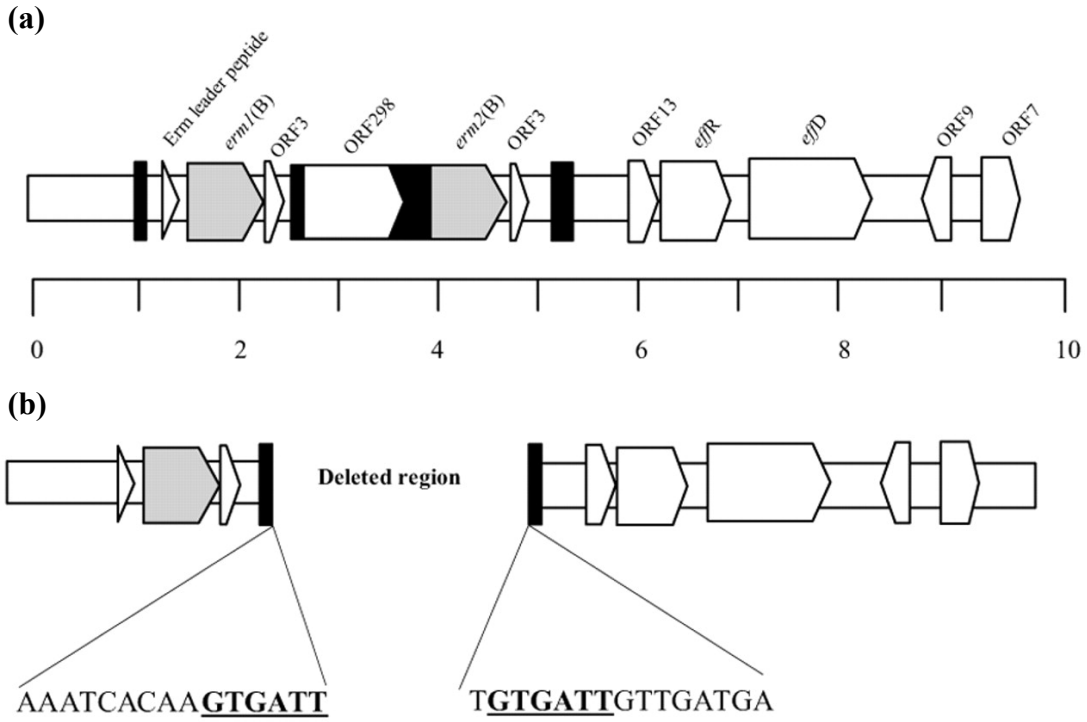



**Figure 4.3:** Genomic structure of *Tn5398* within the *C. difficile* chromosome prior to and following spontaneous deletion of *erm2*(B) gene. (A) The structure of the mobile genetic element *Tn5398*. (B) The structure of *Tn5398Δerm* in which Orf298, Orf3 and *erm*2(B) genes have been deleted. The directions of the arrowed boxes refer to the Open Reading Frames (ORFs) (Hussain *et al.,* 2005).

Primers dnaKF/dnaKR (Jain *et al.,* 2017) were used to confirm the presence of the *dnaK* gene, and as expected they amplified a 210 bp product in all strains apart from the *C. difficile dnaK* mutant in which they amplified a 2059 bp product signifying the presence of the group II intron in correct position within the *C. difficile* 630 *dnaK* chromosome (Jain *et al*., 2017). To confirm the presence of the *spo0A* mutant strain the primer set spo0AF/spo0AR (Heap *et al*., 2007) amplified a 300 bp product in all of the strains apart from the *spo0A* mutant, where amplification with these primers yielded a 2300 bp product signifying, again, the insertion of the group II intron in the *C. difficile spo0A* mutant chromosome thereby confirming the presence of the insertion

within the strain. The EBS universal primer in conjunction with the target gene reverse primer (either *spo0A*-R (500 bp) or *dnaK*-R (300 bp)) amplifies across the intron-exon junction in the antisense orientation, thereby confirming the intron is inserted within the target gene in the correct position (Figure 4.1; Figure 4.4c and 4.4d) (Heap *et al*., 2007; Jain *et al*., 2017).

**Table 4.2:** Expected amplicon size in base pairs (bp).

| | CD630 | CD630*Δerm* | CD630 *dnaK* | CD630 *spo0A* |
|---|---|---|---|---|
| **4140/4620** | 480bp | 480bp | 480bp | 480bp |
| **4140/5880** | 1740 bp | No product expected | No product expected | No product expected |
| **5880/7020** | 1140 bp | No product expected | No product expected | No product expected |
| **dnakF/dnakR** | 210 bp | 210 bp | 2059 bp | 210 bp |
| **spo0AF/spo0AR** | 300 bp | 300 bp | 300 bp | 2300 bp |
| **EBS universal/dnaKR** | No product expected | No product expected | 300 bp | No product expected |
| **EBS universal/spo0AR** | No product expected | No product expected | No product expected | 500 bp |

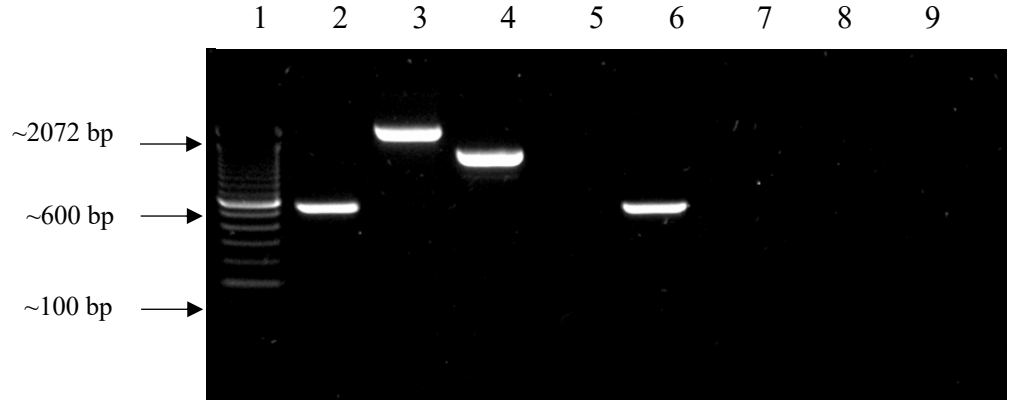



**Figure 4.4a: Genotypic discrimination between 630 and *Δerm* variants.** Gel electrophoresis image of PCR amplicons generated with *Clostridioides difficile* genomic DNA from strains 630 and *Δerm*. **Lane 1**: 100 bp molecular weight marker. **Lane 2**:630 + 4140F/4620R (480 bp); **Lane 3**: 630 + 4140F/5880R (1740 bp); **Lane 4**: 630 + 5880F/7020R (1140 bp); **Lane 5**: negative control 4140F/4620R; **Lane 6**: *Δerm* + 4140F/4620R (480 bp); **Lane 7**: *Δerm* + 4140F/5880R (no band); **Lane 8**: *Δerm* + 5880F/7020R (no band); **Lane 9**:negative control 4140F/5880R.

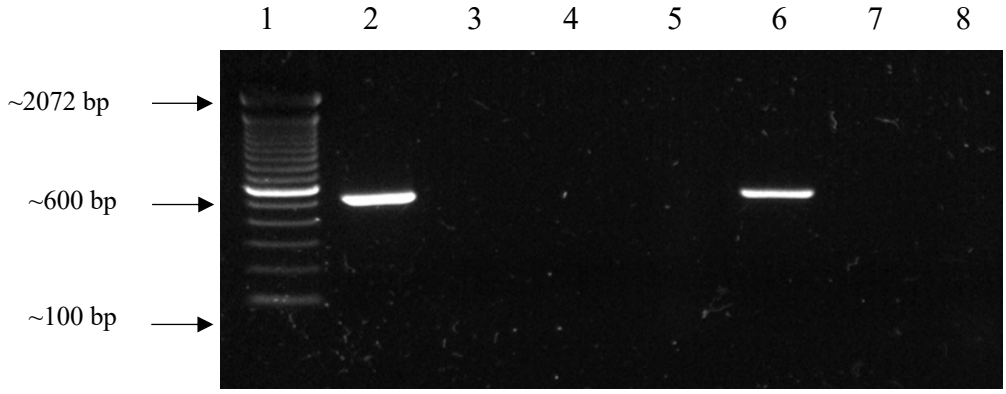



**Figure 4.4b: Genotypic discrimination between 630 and *Δerm* variants**. Gel electrophoresis image of PCR amplicons generated with *Clostridioides difficile* genomic DNA from mutant strains *dnaK* and *spo0A*. **Lane 1**: 100 bp molecular weight marker. **Lane 2**: dnaK mutant + 4140F/4620R (480 bp); **Lane 3**: dnaK mutant + 4140F/5880R (no band); **Lane 4**: dnaK mutant + 5880F/7020R (no band); **Lane 5**: Negative control 5880F/7020R; **Lane 6**: spo0A mutant + 4140F/4620R (480 bp); **Lane 7**: spo0A mutant + 4140F/5880R (no band); **Lane 8**: spo0A mutant + 5880F/7020R (no band).

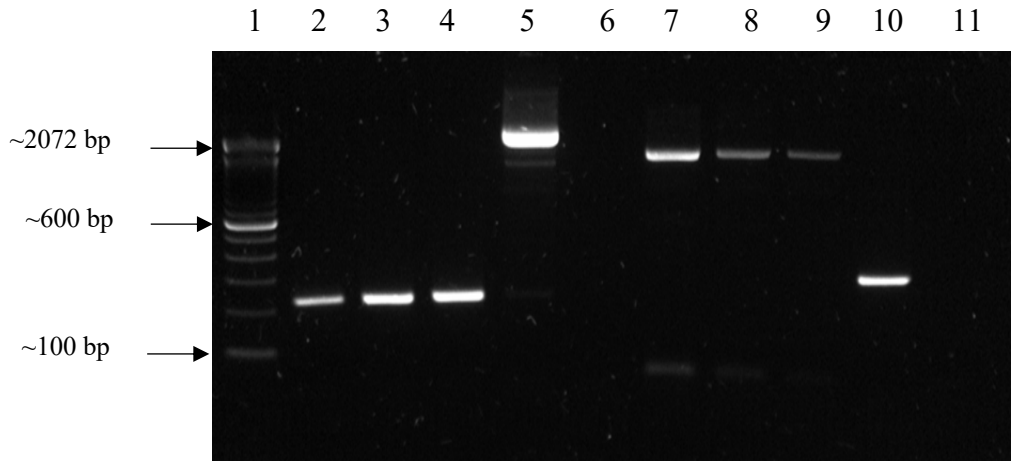



**Figure 4.4c: Genotypic discrimination between 630 and *Δerm* variants**. Gel electrophoresis image of PCR amplicons generated with *Clostridioides difficile* genomic DNA from strains 630, *Δerm*, *dnaK* mutant and *spo0A* mutant. **Lane 1**: 100 bp molecular weight marker. **Lane 2**: 630 + dnaKF/dnaKR (210 bp); **Lane 3**: *Δerm* + dnaKF/dnaKR (210 bp); **Lane 4**: *spo0A* mutant + dnaKF/dnaKR (210 bp); **Lane 5**: *dnaK* mutant + dnaKF/dnaKR (2059 bp); **Lane 6**: Negative control dnaKF/dnaKR; **Lane 7**: 630 + EBS universal/dnaKR (1000 bp); **Lane 8**: *Δerm* + EBS universal/dnaKR (1000 bp). **Lane 9**: *spo0A* mutant + EBS universal/dnaKR (1000 bp). **Lane 10**: *dnaK* mutant + EBS universal/dnaKR (300 bp). **Lane 11**: Negative control + EBS universal/dnaKR.

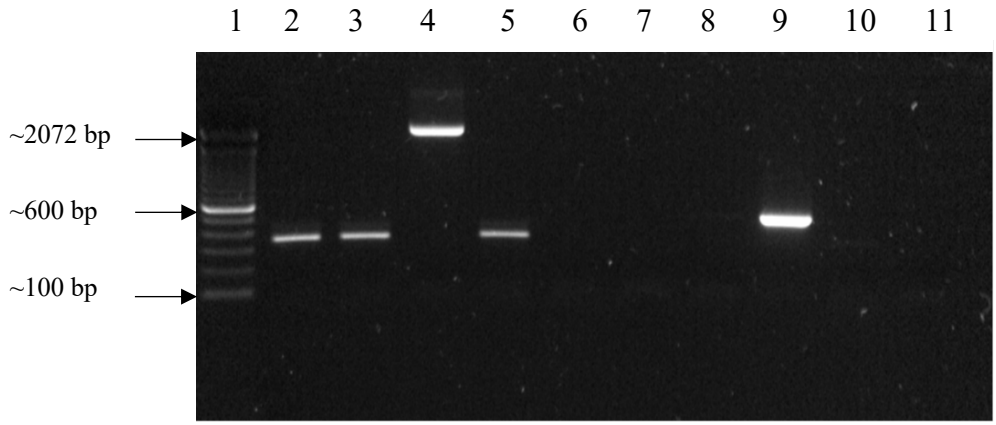



**Figure 4.4d: Genotypic discrimination between 630 and *Δerm* variants**. Gel electrophoresis image of PCR amplicons generated with *Clostridioides difficile* genomic DNA from strains 630, *Δerm*, *dnaK* mutant and *spo0A* mutant. **Lane 1**: 100 bp molecular weight marker. **Lane 2**: 630 + spo0AF/spo0AR (300 bp); **Lane 3**: *Δerm* + spo0AF/spo0AR (300 bp); **Lane 4**: *spo0A* mutant + spo0AF/spo0AR (2000 bp); **Lane 5**: *dnaK* mutant + spo0AF/spo0AR (300 bp); **Lane 6**: Negative control spo0AF/spo0AR; **Lane 7**: 630 + EBS universal/spo0AR (no band); **Lane 8**: *Δerm* + EBS universal/spo0AR (no band). **Lane 9**: *spo0A* mutant + EBS universal/spo0AR (~500 bp). **Lane 10**: *dnaK* mutant + EBS universal/spo0AR (no band). **Lane 11**: Negative control + EBS universal/spo0AR.

As genotyping PCR successfully confirmed the identity of each strain, I was able to proceed with all subsequent work assured that the correct *C. difficile* strain was being used.

### 4.3.3 Growth in FW media

In order to study the growth rates of the four *Clostridioides difficile* strains – 630 (wildtype), *Δerm* (parental), *dnaK* (mutant) and *spo0A* (mutant) in a more biologically

representative environment, the strains were grown in BHIS/faecal water (FW) (50/50 v/v) with BHIS/PBS (50/50 v/v) as a control. Attenuance (D650 nm) was measured in triplicate over a time period of six to seven hours depending on strain and rate of growth. The growth rate constant (ƙ) and doubling time ($T_d$) was calculated for each strain grown in each growth medium from attenuance values collected hourly. Exponential growth measurements from the linear part of the growth curve were used as outlined below (Neidhardt *et al*., 1990) (Table 4.3 and Figure 4.6).

**Formula used to calculate growth rate constant (ƙ):** $ƙ = (\log^{10}N - \log^{10}N0) \times 2.303/ (t\text{-}t_0)$

$N_0 = OD_{650}$nm at time 0 h

$N = OD_{650}$nm at the time of exponential growth ending
t = time of exponential growth ending (h)
$t_0$ = time at inoculation (0 h)

**Formula used to calculate doubling time ($T_d$):** $T_d = 0.693 / ƙ$

*C. difficile* 630 displayed a slightly quicker doubling time in the FW medium compared to the PBS control, 97 min and 104 min respectively; there was no significant difference (Anova, p-value > 0.05) between the two growth conditions at any timepoint in this strain (Figure 4.5a). The *Δerm* strain exhibited the quickest doubling time of all four strains with times of 74 min in PBS and 85 min in FW, when analysed there was a significant difference in attenuance values in the FW medium at timepoint 5 (Figure 4.5b) (Anova, p-value < 0.05). However, this difference was not observed at timepoint 6 or 7 (Figure 4.5b). The *dnaK* mutant displayed the longest doubling time of all strains (139 min in PBS/452 min in FW). Attenuance values were lower the *dnak* mutant strain grown in the FW medium throughout the growth curve

with significant differences observed at timepoints 4, 5 and 6 (Figure 4.5c) (Anova, p-value < 0.05). Doubling time in the *spo0A* mutant was again slower in the FW medium at 95 min in PBS and 160 min in FW but there was no significant difference (Anova, p-value > 0.05) in attenuance at any timepoint (Figure 4.5d). Overall, FW appears to reduce growth rate in the mutant strains but increase growth rate in the wildtype *C. difficile* 630 strain (Table 4.3 and Figure 4.6).

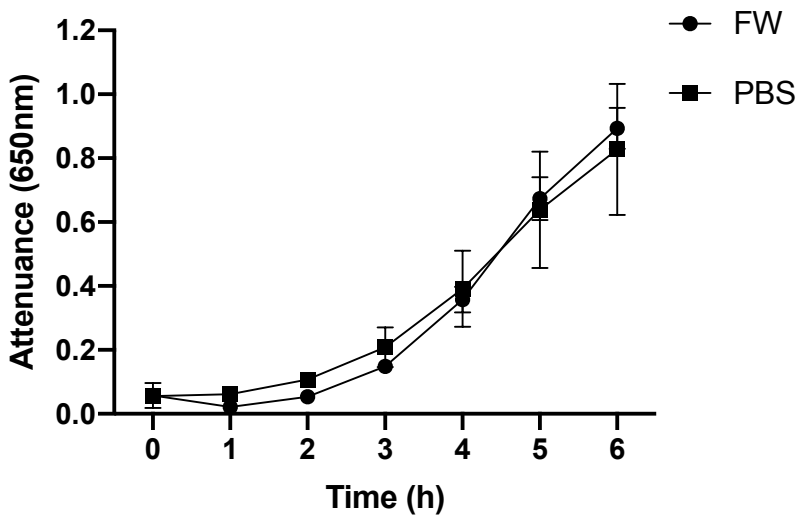



**Figure 4.5a:** Growth of *Clostridioides difficile* 630 in BHIS/faecal water (FW) (50/50 v/v) and BHIS/PBS (50/50 v/v) media. Data presented are means of three independent biological replicates and error bars represent the standard deviation of the mean. Anova showed no significant difference between the two growth conditions.

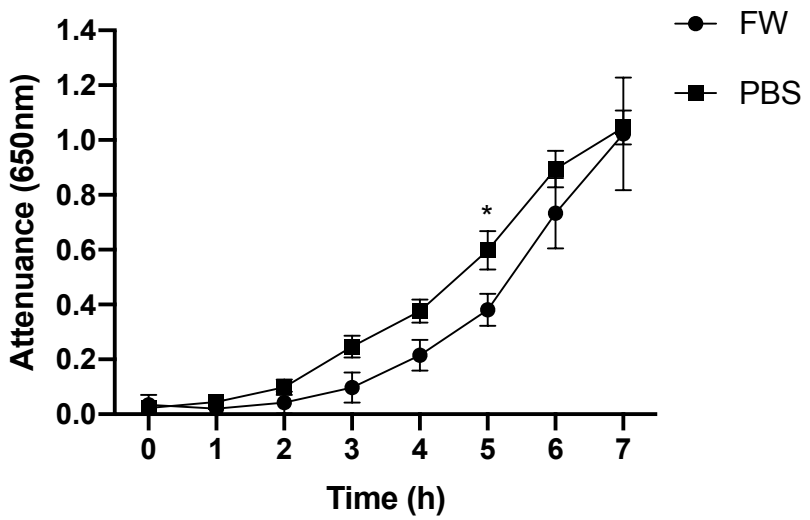



**Figure 4.5b:** Growth of *Clostridioides difficile Δerm* in BHIS/faecal water (FW) (50/50 v/v) and BHIS/PBS (50/50 v/v) media. Data presented are means of three independent biological replicates and error bars represent the standard deviation of the mean. *p < 0.05, Anova, Post hoc: Bonferroni.

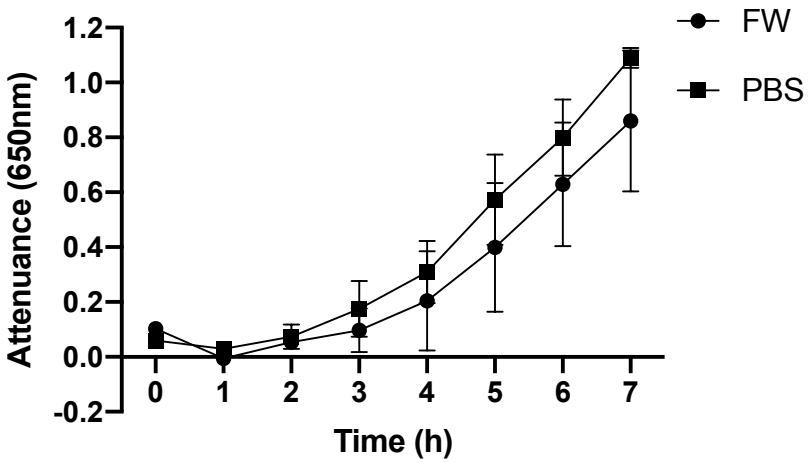



**Figure 4.5c:** Growth of *Clostridioides difficile spo0A* in BHIS/faecal water (FW) (50/50 v/v) and BHIS/PBS (50/50 v/v) media. Data presented are means of three independent biological replicates and error bars represent the standard deviation of the mean. Anova showed no significant difference between the two growth conditions.

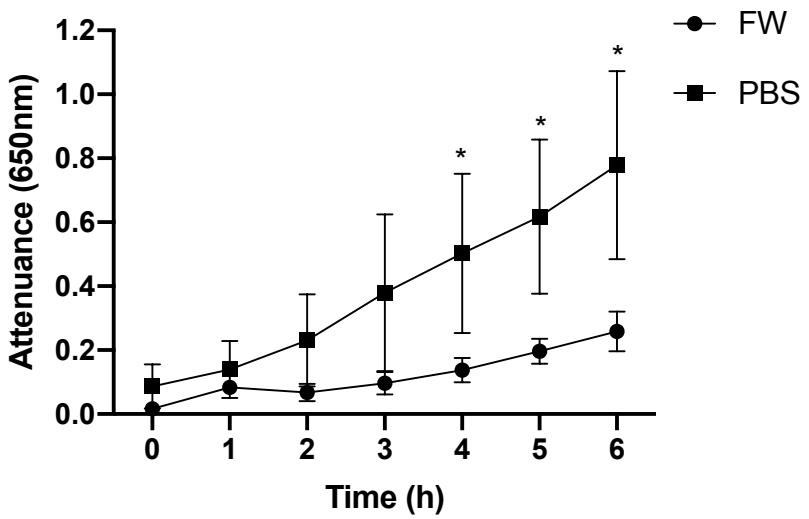



**Figure 4.5d:** Growth of *Clostridioides difficile dnaK* mutant strain in BHIS/faecal water (FW) (50/50 v/v) and BHIS/PBS (50/50 v/v) media. Data presented are means of three independent biological replicates and error bars represent the standard deviation of the mean. *p < 0.05, Anova, Post hoc: Bonferroni.

**Table 4.3:** Mean growth rate constant, ƙ (hr-1) and the mean doubling time, Td (min) for all four strains determined from linear section of growth curve.

| **Growth Medium** | 630 | | *Δerm* | | *dnaK* | | *spo0A* | |
|---|---|---|---|---|---|---|---|---|
| | ƙ ($h^{-1}$) | Td(min) | ƙ ($h^{-1}$) | Td(min) | ƙ ($h^{-1}$) | Td(min) | ƙ ($h^{-1}$) | Td(min) |
| **PBS** | 0.40 | 104 | 0.56 | 74 | 0.30 | 139 | 0.44 | 95 |
| **FW** | 0.43 | 97 | 0.49 | 85 | 0.09 | **452** | 0.26 | 160 |

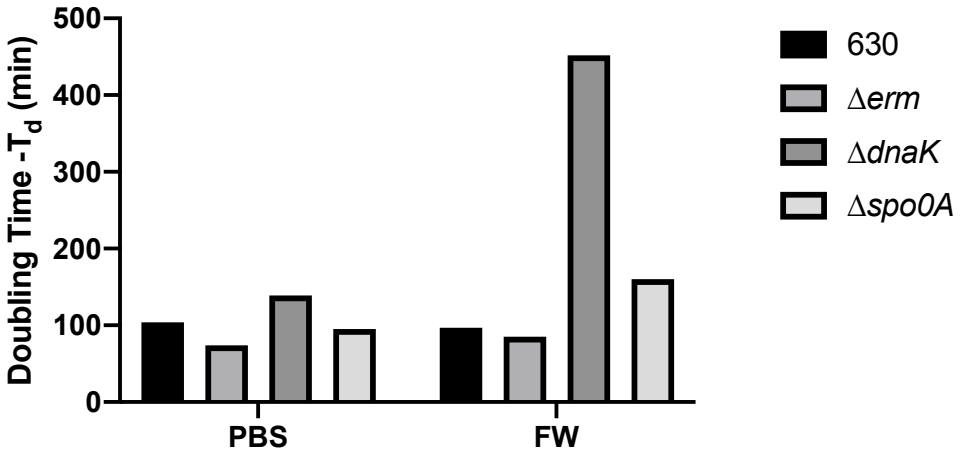



**Figure 4.6:** Mean doubling time of the *Clostridioides difficile* strains 630, *Δerm*, *dnaK* and *spo0A* grown in BHIS/faecal water (FW) (50/50 v/v) and BHIS/PBS (50/50 v/v) media.

### 4.3.4 Cell length

Previous research has shown that, in certain bacteria, cell length increases in response to stress (Justice *et al*., 2004; Pavlostathis *et al*., 2006; Justice *et al*., 2008; Alotaibi, 2017; Jain *et al*., 2017), in addition to this it has been hypothesised that the *dnaK* mutant strain is in a permanent state of stress and displays a significantly longer cell morphology compared to the wildtype (Alotaibi, 2017; Jain *et al*., 2017). Therefore, in order to further characterise the effect of faecal water metabolites on each of the three strains, samples collected at each timepoint were visualised using microscopy (see Chapter 2 for microscope and software used) and the cell length measured; 100 cells were measured per sample. We found that in the *C. difficile* 630 *Δerm* strain there was a significant increase in cell length in the FW grown cells at T1 and T2 with an average length of 5.36 and 5.29 μm when compared to the T0 control (4.40 μm). Cell

length in the *dnaK* mutant strain was shown to be significantly increased from T2 onwards until T7 in the FW samples (7.12, 7.66, 7.48, 7.60, 7.27 µm respectively) when compared to T0 (5.97 µm). In the *spo0A* mutant a significant increase in cell length was observed in the FW samples at T2 (4.85 µm) and T7 (4.92 µm) only (T0 = 4.0 µm) (Figures 4.7a, b, c and d). This increase in length in the FW grown cells mirrors the results found for the wildtype *C. difficile* 630 strain (Chapter 3, Figure 3.13) and shows that FW has a substantial effect on cell morphology.

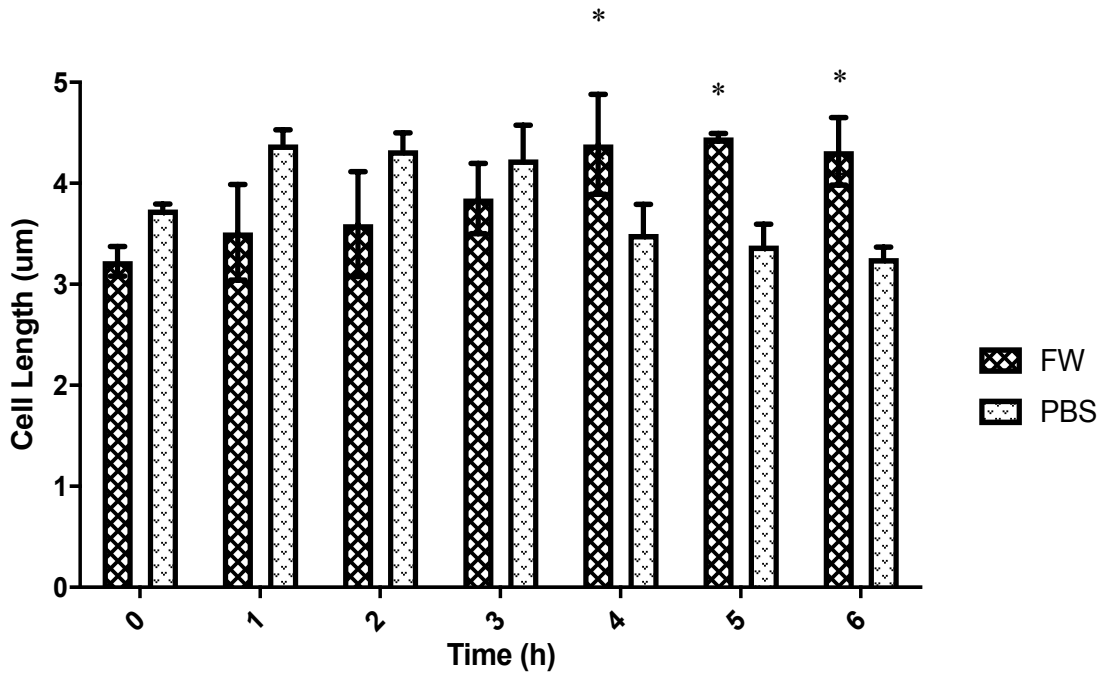



**Figure 4.7a:** Cell length of *C. difficile* 630 grown in BHIS/faecal water (50/50v), and BHIS/PBS (50/50v). Cells are significantly longer during log phase of growth in FW media. A total of 100 cells were measured per sample. The data presented are the means of 3 biologically independent experiments and error bars represent standard deviation of the mean. *p < 0.05, Anova, Post hoc: Dunnett t (2-sided).

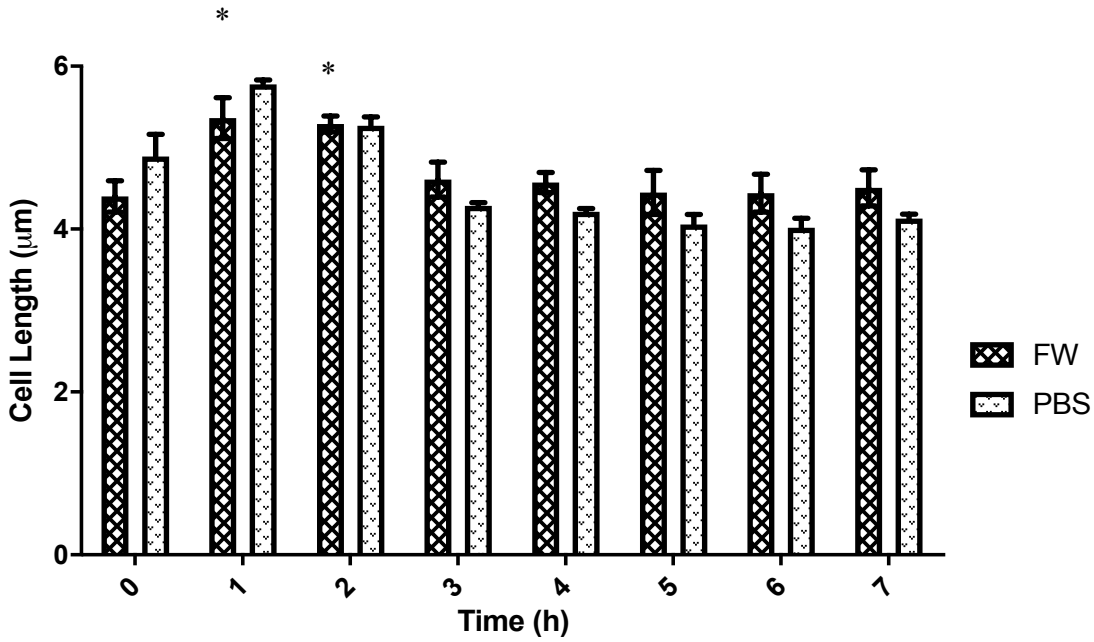



**Figure 4.7b:** Cell length of the parental strain *Δerm* grown in BHIS/faecal water (50/50v) and BHIS/PBS (50/50v). Cells are significantly longer during early-log phase of growth in FW media. A total of 100 cells measured per sample. Means of 3 independent experiments. Error bars represent standard deviation of the mean. *$p < 0.05$, FW = Anova, Post hoc: Dunnett t (2-sided); PBS = Mann Whitney U.

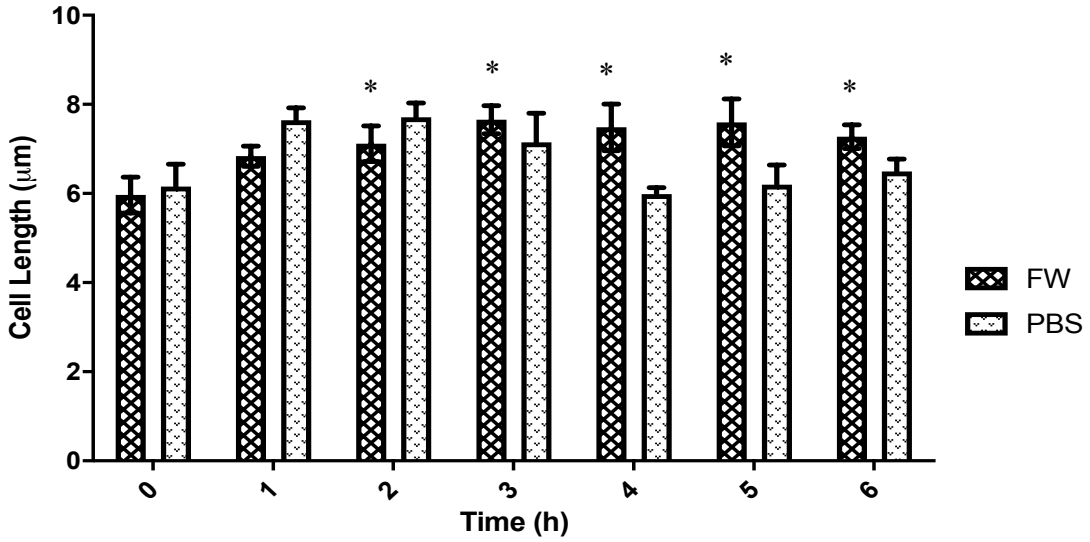



**Figure 4.7c:** Cell length of the *dnaK* mutant strain grown in BHIS/faecal water (50/50v) and BHIS/PBS (50/50v). Cells are significantly longer from early-log phase onwards in the FW media. A total of 100 cells measured per sample. Means of 3 independent experiments. Error bars represent standard deviation of the mean. *p < 0.05, Anova, Post hoc: Dunnett t (2-sided).

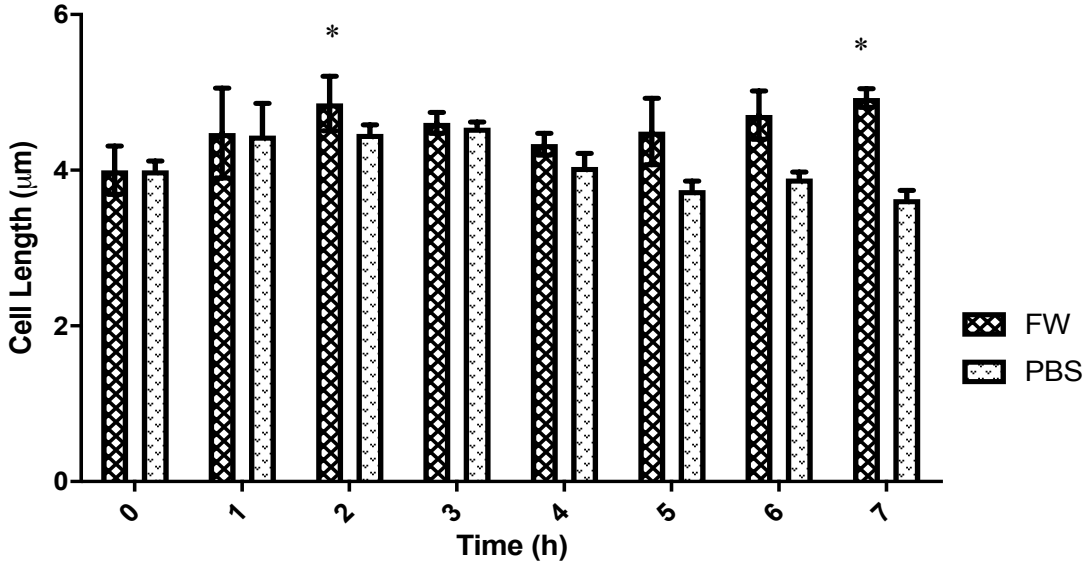



**Figure 4.7d:** Cell length of the *spo0A* mutant strain grown in BHIS/faecal water (50/50v) and BHIS/PBS (50/50v). Cells are significantly longer during early-log phase and late-log phase of growth in FW media. A total of 100 cells measured per sample. Means of 3 independent experiments. Error bars represent standard deviation of the mean. *p < 0.05, Anova, Post hoc: FW = Dunnett t (2-sided); PBS = Dunnett t3.

In order to determine whether there was any correlation between the cell length changes observed and sporulation within the strains, spore staining was undertaken to attempt to enumerate the spores present at each growth point (10 fields of view per sample). Figure 4.8a and b show representative spore stains from *dnaK* and *Δerm* respectively, but following statistical analysis no significant differences were observed in the number of spores per field of view at any timepoint when compared to T0 (Figures 4.9a and b). No spores were identified in mutant *spo0A* as expected. These findings again reflect those observed with *C. difficile* 630 (Chapter 3, Figure 3.12) where no significant change was observed. It was noted that the error bars on Figures 4.9a and b were unusually large (as seen in Figure 3.12), again this may be due to insufficient sample volume or inadequate mixing of sample prior to pipetting onto the glass slide. To avoid this in future I would increase the volume of sample and pipette to mix thoroughly to minimise clumping of spores before adding to the slide.

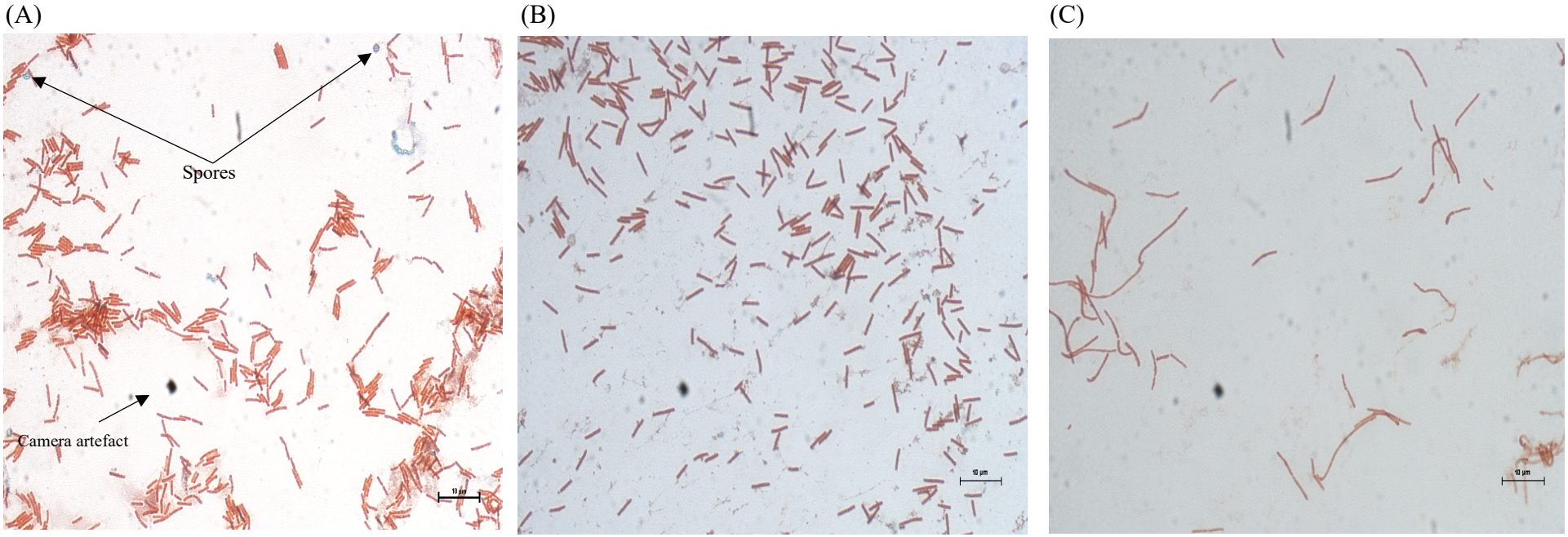



**Figure 4.8:** *Clostridioides difficile* (A) 630, (B) *dnaK* and (C) *Δerm* grown in FW and harvested at T6. Endospore staining carried out with malachite green as described in Chapter 2. Image shows filamentous morphology of the *dnaK* mutant in comparison to the wildtype *Δerm*. Scale bar = 10 µm. Original objective magnification of 1000 X. Nikon Eclipse E400 microscope used equipped with DS-5M-L1 Digital Sight Camera and NIS-Elements Basic Research 3.2 software.

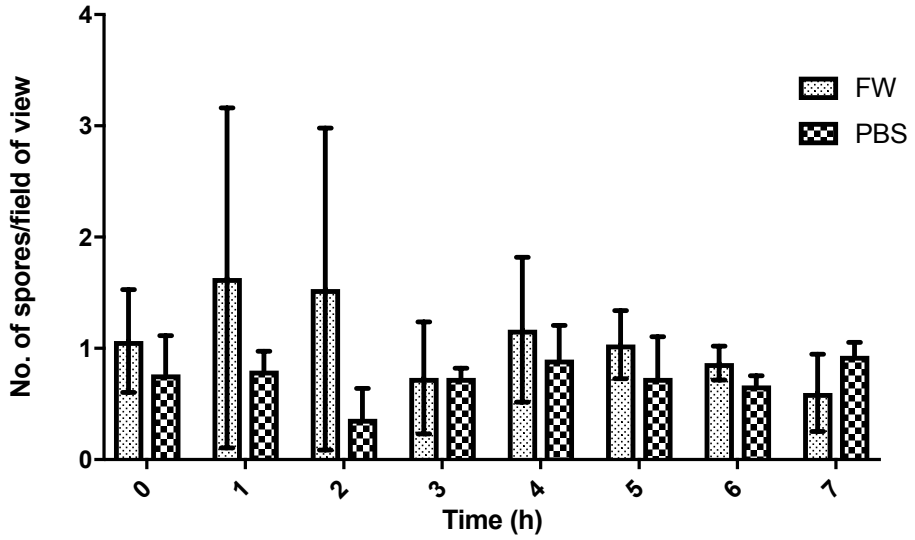



**Figure 4.9a:** No. of spores in *C. difficile Δerm* grown in BHIS/faecal water (50/50v) and BHIS/PBS (50/50v). 10 fields of view per sample. Means of 3 independent experiments. Error bars represent standard deviation of the mean. Anova showed no significant change.

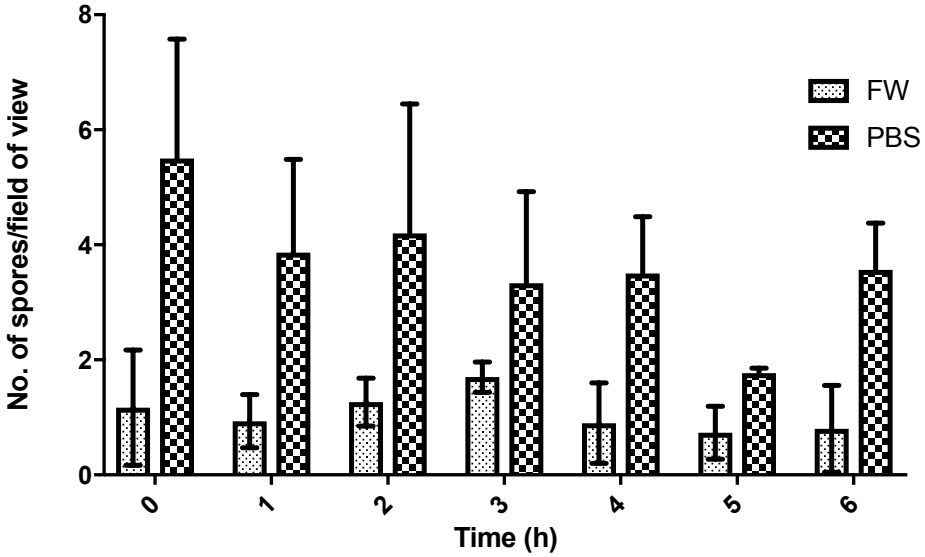



**Figure 4.9b:** No. of spores in *C. difficile dnaK* mutant strain grown in BHIS/faecal water (50/50v) and BHIS/PBS (50/50v). 10 fields of view per sample. Means of 3 independent experiments. Error bars represent standard deviation of the mean. FW = Anova – no significant change; PBS = Kruskal-Wallis – no significant change.

### 4.4.5 Transcriptional analysis of selected genes

Prior to RT-qPCR RNA was extracted from each of the four strains and Bioanalysis (Table 4.4) showed that all *C. difficile* 630 samples had a RIN of ≥ 9, *Δerm* ≥ 9.7, *dnaK* mutant ≥ 7.9 and *spo0A* mutant ≥ 8.8. Based upon the *C. difficile* 630 RNAseq and RT-qPCR data presented earlier (Chapter 3) we know that a number of genes involved in motility, protein folding and toxin production were significantly altered in response to FW in *C. difficile* 630. These observations, combined with measurements showing significantly elongated cells in all four strains grown in FW (Figures 4.7a, b, c and d) led us to further investigate the effect of FW on the transcriptional status of the other three strains. The expression of a number of these genes was assessed by RT-qPCR analysis of the strains *C. difficile* 630, *Δerm*, *dnaK* mutant and *spo0A* mutant, using RNA extracted and reverse transcribed as described in Chapter 3. The molecular chaperones *groEL* and *groES*, motility genes *fliC* and *fliD*, and the gene encoding toxin A (*tcdA*) in *C. difficile* were analysed, using *rpsJ*, *gyrA* and *adK* as reference genes. From this it was found that the expressional fold change for *groEL* (630 – 1.5-fold down; *Δerm* – 2.2-fold up; *dnaK* mutant – 2.8-fold up; s*po0A* mutant – 4.5-fold down) and *groES* (630 – 1.5-fold down; *Δerm* – 1.9-fold up in FW; *dnaK* mutant – 2.4-fold up; *spo0A* mutant – 2.7-fold down) was increased by relatively similar amounts in the *dnaK* mutant and *Δerm* in the FW grown samples, which was in contrast to the strains 630 and *spo0A* mutant where both of these genes exhibited decreased expression.

Motility gene expression showed the greatest variation across the three strains, with *fliC* (flagellin subunit) expression 5-fold down in 630, 58.8-fold down in *Δerm*, 1.1-fold down in *dnaK* mutant and 81.8-fold down in s*po0A* mutant; *fliD* (encoding the

flagellar cap protein) followed a similar pattern, being 4-fold down in 630, 37.3-fold down in *Δerm*, 3.8-fold down in *dnaK* mutant and 38.4-fold down in *spo0A* mutant. Toxin A (*tcdA*) gene expression was decreased by 9.8-fold in 630, 22.4-fold in *Δerm*, 3.7-fold in *dnaK* mutant and 33.7-fold in *spo0A* mutant (Figure 4.10).

**Table 4.4:** Bioanalysis of RNA extraction from *Clostridioides difficile* 630, *Δerm, dnaK* mutant and *spo0A* mutant grown in BHIS containing 50% faecal water (FW), BHIS containing 50% PBS and standard BHIS made with deionized $H_2O$. Three biological replicates were assessed for each growth condition. P=PBS; F=FW. *As the concentration of RNA was low in sample *spo0A* P2 no bioanalysis was carried in order to preserve enough sample for cDNA synthesis.

| RNA Sample | Concentration (ng/µl) | A260/280 | A260/230 | RIN |
|---|---|---|---|---|
| **630 P1** | 1477.6 | 2.19 | 2.38 | 9 |
| **630 P2** | 1523.2 | 2.18 | 2.50 | 9.9 |
| **630 P3** | 1391.6 | 2.18 | 2.41 | 9.8 |
| **630 F1** | 1859.0 | 2.19 | 2.51 | 9.7 |
| **630 F2** | 1643.0 | 2.19 | 2.33 | 9.4 |
| **630 F3** | 1705.5 | 2.19 | 2.52 | 9.7 |
| ***Δerm* P1** | 2009.6 | 2.18 | 2.39 | 10 |
| ***Δerm* P2** | 1571.0 | 2.20 | 2.42 | 10 |
| ***Δerm* P3** | 1472.9 | 2.19 | 2.12 | 10 |
| ***Δerm* F1** | 1346.1 | 2.20 | 2.38 | 10 |
| ***Δerm* F2** | 1615.0 | 2.20 | 2.46 | 9.7 |
| ***Δerm* F3** | 1874.7 | 2.19 | 2.40 | 10 |

| | | | | |
|---|---|---|---|---|
| ***dnaK* P1** | 23.6 | 2.04 | 1.52 | 9 |
| ***dnaK* P2** | 53.0 | 2.06 | 0.61 | 9.5 |
| ***dnaK* P3** | 74.4 | 2.09 | 1.45 | 9.1 |
| ***dnaK* F1** | 53.4 | 1.99 | 1.14 | 7.9 |
| ***dnaK* F2** | 23.2 | 2.05 | 0.69 | 8.2 |
| ***dnaK* F3** | 29.1 | 2.01 | 0.80 | 8.2 |
| ***spo0A* P1** | 76.4 | 2.11 | 1.00 | 9.7 |
| ***spo0A* P2*** | 16.0 | 1.99 | 0.23 | - |
| ***spo0A* P3** | 84.7 | 2.09 | 0.89 | 9.5 |
| ***spo0A* F1** | 177.6 | 2.16 | 1.31 | 8.9 |
| ***spo0A* F2** | 125.5 | 2.32 | 1.08 | 8.8 |
| ***spo0A* F3** | 222.7 | 2.16 | 1.46 | 9.2 |

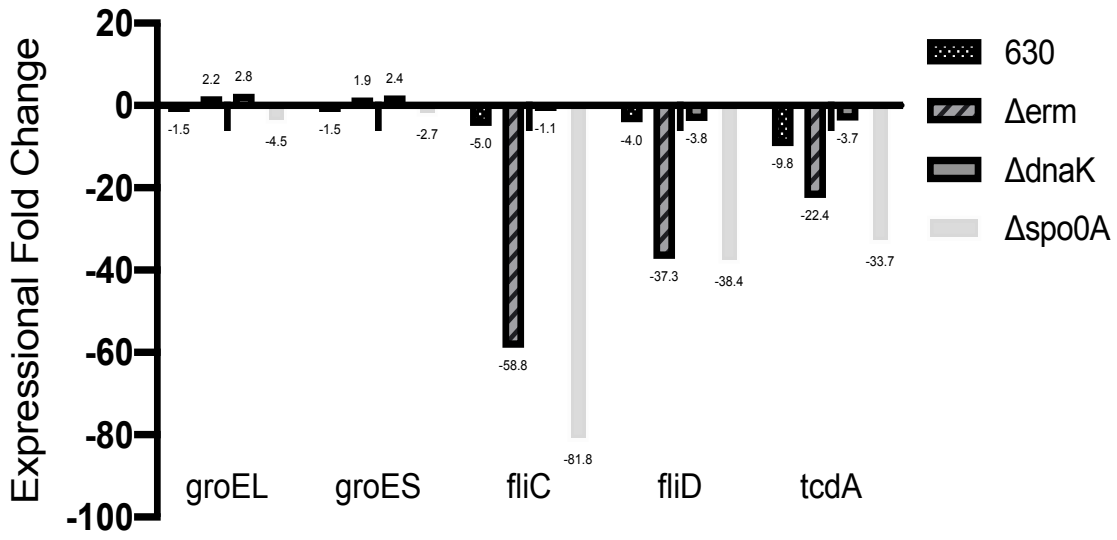



**Figure 4.10:** Comparison of expression of selected genes between *C. difficile* 630, *Δerm*, *dnaK* mutant and *spo0A* mutant gown in FW media. Expressional changes determined by RT-qPCR are shown relative to the BHIS control. *rpsJ*, *gyrA* and *adk* were used as reference genes.

Sporulation gene expression was also investigated in *Δerm* and the *dnaK* mutant as previous observations in *C. difficile* 630 (Chapter 3) of significant changes in the expression of a number of sporulation-associated genes in the FW grown cells led to the hypothesis that such changes could be a response to the presence of FW. The expression of a gene encoding a stage III sporulation protein (*spoIIIAC*) was increased in 630, *Δerm* and the *dnaK* mutant by 161.2, 7.6-fold and 8.9-fold respectively. Transcript levels for *spoVB*, encoding a stage 5 sporulation protein, were found to be down in 630 by 1.1-fold, up in *Δerm* by 1.2-fold, and down in the *dnaK* mutant by 3.1-fold in the FW media. The gene *cspC*, encoding a germination-associated protein, was increased by 38.1-fold in 630, 3.6-fold in *Δerm* and decreased by 1.1-fold in the *dnaK* mutant (Figure 4.11).

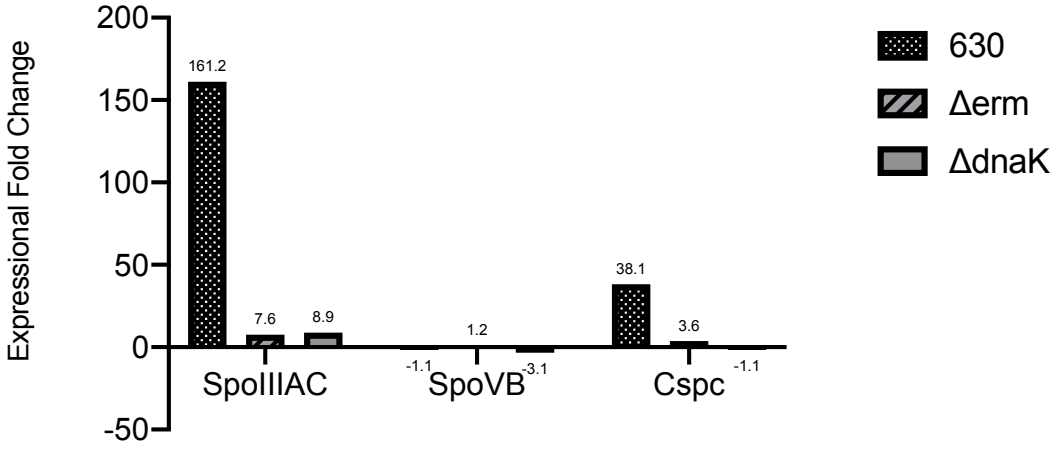



**Figure 4.11:** Comparison between *C. difficile* 630, *Δerm* and *dnaK* mutant of selected sporulation-associated genes. Expressional changes determined by RT-qPCR are shown relative to the BHIS control. *rpsJ*, *gyrA* and *adk* were used as reference genes.

## 4.4 DISCUSSION

### 4.4.1 Growth in FW

We previously found that, for *C. difficile* 630, growth in the presence of FW did not affect growth rate (Chapter 3). We were therefore able to replicate this finding in this set of growth curves as a control measure (Figure 4.5a). However, the mutant strains *dnaK* and *spo0A* exhibited reduced growth rate in the faecal water compared to the parental *Δerm* strain (Figures 4.5a, b, c and d). There was also much greater variation between biological replicates observed in the *dnaK* mutant grown in the BHIS media highlighting the variability of growth in this strain in particular. Jain *et al.* (2017) previously observed a reduction in growth rate of the *dnaK* mutant strain in BHIS media at 37 °C compared to the parental *Δerm* and suggested that this may be due to a deficiency in the cellular synthesis of DNA and RNA as previously observed by Itikawa and Ryu (1979) in *E. coli dnaK* mutants. Reduced growth rates and lowered stress tolerance was also observed in a *C. botulinum dnaK* ClosTron mutant by Selby *et al*., (2011). Further reduced growth of the *dnaK* mutant in the FW media may suggest that FW creates a stressful environment for the mutant organism which is already lacking, we hypothesise, full ability to control cellular homeostasis and protein folding. Alotaibi (2017) report significantly reduced growth kinetics in a *C. difficile dnaK* mutant strain grown in the presence of p-cresol in comparison to the wild-type. We could therefore hypothesis that p-cresol present in the FW has a bacteriostatic effect on the *dnaK* mutant. Growth of the *spo0A* mutant was similar to the wildtype in the control media, this is in keeping with the growth characteristics seen by Deakin *et al*. (2012) and Pettit *et al*. (2014) where both strains displayed similar growth kinetics. However, reduced growth of the *spo0A* mutant was observed in the FW media, again this may be due to the FW imposing stress on the organism. This may suggest that the

lack of *spo0A* transcription initiation has wider knock-on effects not previously studied when the organism is exposed to FW and therefore opens the door for further investigation into sporulation initiation in the human intestinal environment.

### 4.4.2 Cell length

An increase in cell length was observed in *Δerm, dnaK* mutant and *spo0A* mutant grown in FW media during early-log phase of growth (T2) this may reflect the initial increase in growth and therefore cell divison at this stage of growth. In addition to this we have witnessed a further increase in cell length in the FW media grown *dnaK* mutant cells from T2 onwards. It has been shown previously by Jain *et al*. (2017) that *dnaK* mutant cells are longer (6.85 ± 1.55 μm) in comparision to the wildtype (4.85 ± 1.04 μm). Alotaibi (2017) also observed an increase in cell length in the *dnaK* mutant revealing an average cell length of 11.28 ± 3.6 μm compared to the wildtype *C.difficile* 630 strain at 4.43 ± 1.19 μm. This increase in cell length has been suggested to result from defective cell division within *dnaK* mutant strains (McCarty and Walker, 1994; Schulz *et al*., 1995). Results reported by Sugimoto *et al*. (2008) found that in an *E. coli dnaK* deletion mutant, the filamentous morphology observed was linked to increases in the amount of GrpE, a co-chaperone of DnaK and DnaJ (dnaJK–GrpE operon), which negatively impacted the assembly of the DnaK chaperone system leading to faulty cell division through irregular FtsZ (a cytoskeleton protein that defines the plane of cell division (Margolin, 2005)) localisation (Sugimoto *et al*., 2008). *GrpE* expression has previously been shown to be increased in a C. *difficile dnaK* mutant (grown at 37 ºC) by Jain *et al*. (2017). Furthermore, it has been suggested previously that bacterial filamentation is a possible defense mechanism against host immune

response in *E. coli* (Justice *et al*., 2004). Justice *et al.* (2004) suggest that *E. coli* in mice bladder cells become filamentous as a defence mechanism to host immune responses, by helping them to evade infiltrating polymorphonuclear leukocytes. It is therefore possible that the *dnaK* mutant senses the extra threat from the FW components and utilises filamentation as means to escape potentially toxic compenents within the FW such p-cresol for example (previously shown to affect growth of the *dnaK* mutant by Alotaibi (2017)). The *spo0A* mutant showed an overall cell length of 4.39 ± 0.89 µm and an increase in length at early and late-log but not at mid-log in the FW media, this late-log increase in cell length was not observed in cultures of the parental *Δerm* strain. We could hypothesis that the cell length increases again at late-log as this would normally be a time of heightened spore production due to the reduction in nutrient availability etc. (Pettit *et al*., 2014). As the *spo0A* mutant is unable to sporulate the cell may instead become further filamented as a survival mechanism as described by Justice *et al*. (2004; 2008). Further investigation would be required to elucidate the mechanism behind this late-log increase in cell length within the *spo0A* mutant.

### 4.4.3 Molecular chaperones

The molecular chaperones *groEL* and *groES* were found to be differentially regulated in the FW media across all four strains again emphasising the regulatory effect FW has on gene expression in *C. difficile*. Both 630 and the *spo0A* mutant showed a decrease in the expression of *groEL* (1.5 and 4.5-fold) and *groES* (1.5 and 2.7-fold) whereas the *Δerm* and the *dnaK* mutant showed increased expression of both molecular chaperones by 2 to 3-fold. This is in keeping with previous work where

*groEL* and *groES* were up-regulated by up to 4-fold the *dnaK* mutant (Jain *et al*., 2017). Research in *C. botulinum* by Selby *et al*. (2011) found an upregulation of *groEL* and *groES* after heat stress at 45 °C. These molecular chaperones function to ensure proper folding of polypeptides into their appropriate final protein conformation within the cell by working together, forming the GroEL/ES chaperonin system which has been referred to as a "protein folding cage" (Georgescauld *et al.,* 2014). It is widely accepted that GroEL is an essential component of the natural stress response within prokaryotes and tends to be induced during stress conditions including sporulation, high temperature and biofilm formation (Hennequin *et al*., 2001a; Arora *et al*., 2017). Our findings here are consistent with this in that the faecal environment appears to have induced a stress-response within the *dnaK* mutant organism. In addition to this, the lack of the master regulator of sporulation, Spo0A, within the *spo0A* mutant strain has led to a decrease in expression of the both of these chaperonins. Research in *Bacillus subtilis*, has shown that Spo0A regulates processes other than sporulation, such as metabolism and efflux pumps (Molle *et al*., 2003; Deakin *et al*., 2012). Prevous RNAseq in a *C. difficile spo0A* mutant strain (Pettit *et al*., 2014) did not find a significant change in the expression of *groEL* or *groES*, we can therefore hypothesise that our observations in the reduction in chaperonin expression here are due to the presence of FW and its affect on Spo0A.

### 4.4.4 Motility

The motility genes *fliC* and *fliD* were both down-regulated in all four strains in the FW media. The wildtype strain 630, *dnaK* mutant and *Δerm* all displayed reduced expression of the genes encoding the flagellin subunit (*fliC*) and the flagellar cap

protein (*fliD*). 630 – 5 (*fliC*) and 4-fold (*fliD*) down; *dnaK* mutant – 1.1 (*fliC*) and 3.8-fold (*fliD*) down; *Δerm* – 58.8 (*fliC*) and 37.3-fold down (*fliD*). Reduced expression of *fliC* has been reported previously Jain *et al*. (2017) where the expression of *fliC* was 4-fold lower in the *dnaK* mutant. This decrease in FliC has also been observed in *B. subtilis* (Schulz *et al*., 1995) and *E. coli* (Shi *et al*., 1992) where *dnaK* inactivation led to non-motile phenotypes. The flagellum has been associated with virulence in *C. difficile* by encouraging adherence to epithelial cells within the colon (Tasteyre *et al*., 2000; Tasteyre *et al*., 2001; Poutanen and Simor, 2004). However, other studies in *C. difficile* 630 have shown that mutations in *fliC* and *fliD* led to an increase in virulence (Dingle *et al*., 2011; Aubry *et al*., 2012; Baban *et al*., 2013). These inconsistencies are thought to be due to differences in the production of flagella between strains of *C. difficile* and the introduction of single nucleotide polymorphisms in *Δerm* and its derivatives (Collery *et al*., 2017). The expression of the motility genes was also substantially downregulated in the *spo0A* mutant by up to 82-fold suggesting that motility is decreased under the FW growth conditions. Our results would strongly suggest that both the *dnaK* and s*po0A* genes play a role in the expression of these motility genes at least in the presence of faecal water. Sporulation and motility have been shown to be related functions within *C. difficile* previously as Pettit *et al*. (2014), for example, reported an increase in expression of motility-associated genes including *fliC* at the mRNA transcript level in the *spo0A* mutant under normal growth conditions. They suggest that this apparent negative regulation of motility factors could be a mechanism used by *C. difficile* to avoid detection by the host defences during sporulation, as FliC is immunogenic and has been shown to elicit an IgG response in hamsters (Bruxelle *et al*., 2018). We, however, find the absence of the master regulator of sporulation significantly decreases the production of motility associated genes in

the presence of FW. These observations further emphasise the complex nature of *C. difficile* and suggest that there are many factors involved in the expression of these important virulence factors. This work highlights the importance of further studies more closely aligned to the human colonic environment.

### 4.4.5 Toxin expression

Toxin A (*tcdA*) gene expression was decreased again in all of the FW grown strains of *C. difficile*. This follows RNAseq and RT-qPCR results (Chapter 3) with 630 where we observed a decrease in *tcdA* expression (by 4-fold) thought to be linked to an observed reduction in butyrate metabolism within the FW leading to a reduction in toxin stimulation. The reduction in *tcdA* expression in the *dnaK* mutant supports previous observations made by Alotaibi (2017) who also found a reduction of up to 4-fold in transcript levels of this particular gene within the *dnaK* mutant. Furthermore, the significant decrease in *tcdA* expression in the *spo0A* mutant strain highlights the role of Spo0A in the expression of virulence factors (Ternan *et al*., 2018) (as discussed previously with regards to motility-associated gene expression). A number of previous studies have also suggested that toxin expression is closely related to sporulation (Underwood *et al*., 2009; Deakin *et al*., 2012; Mackin *et al*., 2013; Pettit *et al*., 2014; Edwards *et al*., 2016) and our results presented here further confirm this and further highlight the influence of FW on toxin A gene expression in this organism. This could be confirmed in future studies through the use of a toxin test on FW grown culture supernatants versus BHIS grown cultures where we would expect to see less toxin A in the FW samples.

### 4.4.6 Sporulation

Sporulation gene expression was assessed in the *dnaK* mutant strain versus the 630 wildtype and *Δerm* parental strains. The sporulation-associated gene *spoIIIAC* was upregulated in all of the strains under FW growth conditions. The 630 strain showed a very significant increase in this gene at 161-fold up whereas *Δerm* and the *dnaK* mutant displayed more modest increases in expression of 7.6 and 8.9-fold, respectively. It could be suggested that unkown mutations introduced in the development of the *Δerm* strain may be responsible for these differences. However, the increases in sporulation in the FW samples does follow the data found in Chapter 3 and further confirms the role of FW in sporulation which is confirmed by the increase in cell length observered. *SpoVB* (involved in stage 5 of sporulation) was reduced slighly (3.1-fold) in the *dnaK* mutant suggesting that the *dnaK* mutant may reach this stage of sporulation at a much later attenuance than the parental strains which would follow the growth pattern observed in this strain. We also found that the gene *cspC* (encoding a germination-associated protein) was 2.5-fold downregulated in the FW grown *dnaK* mutant strain compared to 630 at 38.1-fold up and *Δerm* at 3.6-fold up. This differential expression of the sporulation-associated genes indicates a potential regulatory function of the *dnaK* chaperone with regards to sporulation in the presence of FW within *C. difficile*. Supporting this idea, Li *et al*. (2010) found that the molecular chaperone GroEL1 had a regulatory function in sporulation in the organism *Myxococcus xanthus* – a Gram-negative predatory bacterium capable of forming multicellular structures during sporulation (Kroos *et al*., 1986). The link between molecular chaperones and sporulation in *C. difficile* could be further elucidated with a more in-depth analysis of sporulation within the *dnaK* mutant. Enumeration of the

spores present was attempted by staining and microscopy but no significant data could be obtained from this.

To summarise, this chapter represents the first study into how FW affects the *C. difficile* mutant strains *dnaK* and s*po0A* versus the wildtype strain 630 and parental strain *Δerm.* We have shown that this growth environment, which we believe is physiologically more akin to the human intestine in which *C. difficile* causes disease, results in increased cell length in the mutant strains especially in *dnaK*. We also see that molecular chaperone, motility, toxin and sporulation gene expression is differentially expressed in the presence of FW with clear differences in the behaviour of the mutant strains strongly indicating not only that FW has an effect on *C. difficile* and thereby virulance and pathogenesis promoting capabilities but that the genes *spo0A* and *dnaK* may play important roles in the regulation of these responses. We have therefore provided a more realistic indication of how this organism may respond in the presence of human gut constituents.

# CHAPTER 5: ANALYSIS OF THE EXPRESSION OF SMALL REGULATORY RNAs IN *CLOSTRIDIOIDES DIFFICILE* 630

## 5.1 INTRODUCTION

All organisms express small RNA (sRNA) molecules, the majority of which act to regulate gene expression by altering the secondary structure of the mRNA transcript (Storz *et al*., 2004; Pichon and Felden, 2005). The discovery of sRNA molecules as important regulators of cellular activity in both eukaryotic and prokaryotic organisms has led to an increase in the understanding of RNA function (Garzon *et al.*, 2009). The regulatory function of RNA was confirmed by Fire *et al*. (1998) who discovered RNA interference (RNAi) in the nematode worm, *C. elegans*. The RNAi pathway is found in most eukaryotic organisms as a sequence-specific gene silencing mechanism; this pathway involves the use of the enzyme Dicer, which cuts long sequences of double stranded RNA (dsRNA) into smaller pieces of RNA. These shorter sections of RNA alongside Argonaut proteins form the RNA-Induced Silencing Complex (RISC), which induces mRNA cleavage (Mack, 2007). The target dsRNA can be either exogenous (i.e. arising externally e.g. from infection by a virus) or endogenous (i.e. arising from within the cell e.g. naturally occurring long self-complementary RNA transcripts) (Hannon *et al*., 2002; Putral *et al*., 2006; Valencia-Sanchez *et al.*, 2006). It is believed that RNAi is a cellular defence mechanism, protecting cells from viruses and transposons (Bagasra and Prilliman, 2004; Li *et al*., 2006). This discovery has led to the use of dsRNA as a molecular tool – synthetic dsRNA can be introduced into a cell in order to target a particular gene or set of genes and cause gene silencing from activation of the RNAi pathway (Setten *et al*., 2019). The synthetic dsRNA is complementary to the intended target, once bound to the mRNA of choice (after being made single-stranded by the RISC complex) the cell recognizes this dsRNA as abnormal and degradation occurs, thereby preventing translation of the target gene. This can be useful for research into the function of a particular gene in an organism or

tissue – by silencing a gene of interest the overall effect on the organism can be observed (Bernstein *et al*., 2001; Voorhoeve, and Agami, 2003; Setten *et al*., 2019).

It was initially believed that no such regulatory mechanisms existed in prokaryotes until bacterial sRNAs were first identified over 40 years ago with the discovery of 6s RNA in *E. coli* (Vogel and Wagner, 2007; Vogel, 2009). However, it is only recently that their importance in the regulation of the bacterial genome and their influence in many cellular processes has become apparent. This insight has been gained through systematic genome-wide investigations, both *in silico* and laboratory based with the use of microarrays, PCR, blotting and sequencing techniques (Masse and Gottesman, 2002; Vogel and Wagner, 2007). Bacterial sRNAs are typically between 50 and 500 nucleotides in length and are located either within the intergenic regions (IGRs) (usually *trans*-acting) of the bacterial genome or on the antisense strand (usually *cis*-acting). The majority of the sRNAs in prokaryotes are non-coding RNAs i.e. they do not produce their own protein product. Many of them behave as post-transcriptional regulators by bonding to specific targets (Mulhbacher *et al*., 2010; Chen *et al*., 2011). As with siRNA and miRNAs in eukaryotic organisms, the sRNA is able to either bind directly to protein targets in order to alter the function of the protein or they can bind to the mRNA transcripts to regulate gene expression by modulating translation initiation or target stability (Padalon-Brauch *et al*., 2008; Waters and Storz, 2009).

There are two main types of bacterial sRNA, *cis*-encoded sRNAs and *trans*-encoded sRNAs (Figure 5.1). *Cis*-encoded sRNAs often involve antisense RNAs that overlap the target gene and are fully complementary to their target. *Trans*-encoded sRNAs are

characterised by a sequence that is partially complementary to the target gene RNA, and which is encoded in a separate genomic location to the target gene (Cao *et al*., 2010; Mraheil *et al*., 2010). *Trans* sRNAs are often facilitated by the protein Hfq, which was first discovered in 1968 in *E. coli* as an essential host factor of the RNA bacteriophage Qβ (where the name Hfq originates) (Franze de Fernandez *et al*., 1968). This protein works to increase the stability and fidelity of sRNAs and the interactions between them and their targets, leading to either negative or positive regulation of gene expression by modulating RNA stability or translation (Christiansen *et al*., 2004, Boudry *et al*., 2014). The importance of Hfq in *C. difficile* 630 has been highlighted by the creation of an Hfq-depleted (knockdown mutant) strain (Boudry *et al*., 2014). It was found that the absence of Hfq affected growth, morphology, and increased the organism's sensitivity to stress alongside increased sporulation and biofilm formation. An example of a *trans*-acting sRNA that utilises Hfq is MicC, a 109 nucleotide sRNA, which base-pairs with *ompD* in *S. Typhimurium* causing mRNA degradation by RNase E (Pfeiffer *et al*., 2009).

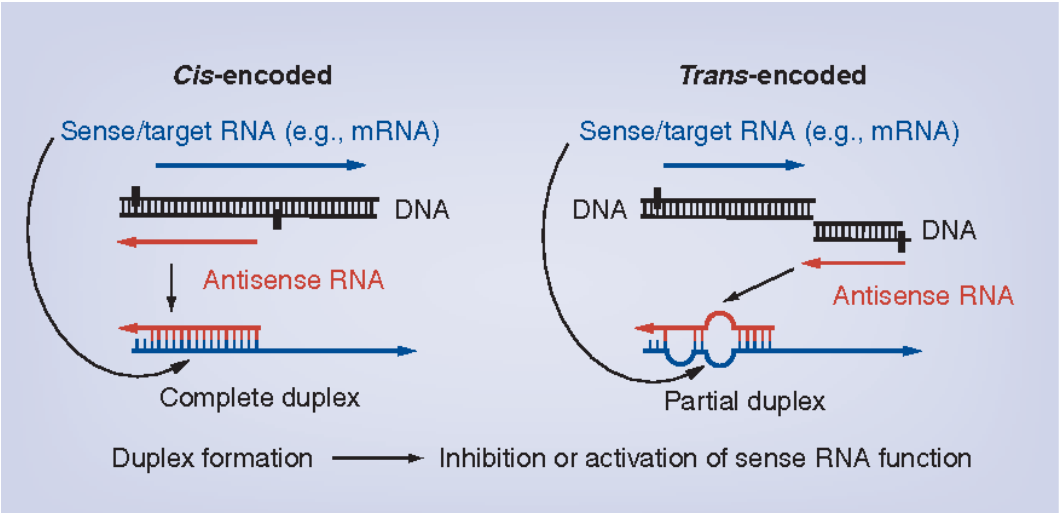



**Figure 5.1:** Schematic showing *cis*- and *trans*-encoded antisense RNAs (Brantl, 2012).

Many bacterial sRNAs behave as riboswitches, which are RNA molecules that are able to detect changes in cellular stimuli such as metabolite or nutrient changes. Riboswitches were first discovered in *E. coli* (Nahvi *et al*., 2002) but have since been found in plants, algae, and yeast suggesting that these molecules provide an important regulatory function in many living organisms (Shimizu *et al*., 2002; Andre *et al*., 2008; Wan *et al.*, 2011). The binding of a metabolite to the aptamer section of a riboswitch results in a conformational change in neighboring mRNA molecules, this affects transcription by either creating a terminator or anti-terminator stem, or translation by either blocking or revealing the Shine-Dalgarno sequence (ribosome binding site located up stream of the start codon AUG) (Figure 5.2) (Barrick and Breaker, 2007; Kim and Breaker, 2008).

It has been shown that sRNAs are involved in stress response, iron homeostasis, outer membrane protein biogenesis, sugar metabolism, quorum sensing and bacterial virulence (Repoila *et al*., 2003; Ziebandt *et al*., 2004; Mraheil *et al*., 2010). Many riboswitches have been found to respond to the signalling molecule cyclic-di-guanosyl-5'monophosphate (c-di-GMP) to regulate various pathogenic processes including, swimming and surface motility, intestinal colonization, biofilm development and toxin secretion by sensing specific receptors that respond by changing their activity or function (McKee *et al*., 2018). A number of the sRNAs in *C. difficile* are associated with membrane proteins that may act as antibiotic transporters or efflux mechanisms to facilitate antibiotic resistance, where microorganisms are able to actively pump toxic substances (such as antibiotics) out of the cell (Soto, 2013). A small RNA has been identified by Chen *et al*. (2011) as a

modulator of gene expression in *C. acetobutylicum* where it functioned as a *cis* acting riboswitch in response to the antibiotic clindamycin. The sRNA, designated sCAC610, is located upstream of an ATP-binding cassette (ABC) transporter gene (CAC0528), both the sRNA and the target gene displayed up to 6.5-fold increases in expression following exposure to the clindamycin, suggesting an involvement in efflux pump activity within this organism (Chen *et al*., 2011).

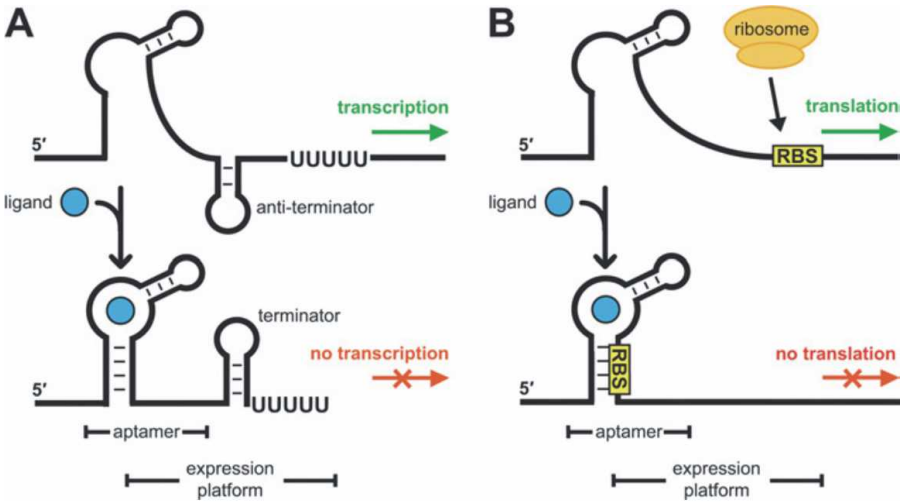



**Figure 5.2:** Methods of riboswitch action. A: premature transcription termination B: inhibition of translation initiation. RBS = Ribosome binding site (Kim and Breaker, 2008).

The functions of a variety of sRNAs in *C. difficile* have been predicted previously (Tables 5.2 and 5.3) (Chen *et al*., 2011; Li *et al.,* 2013; Soutourina *et al.,* 2013). Previous systems biology investigations at Ulster University of proteome and

transcript changes in response to clinically relevant heat stress in *C. difficile* have shown that in some cases, transcript and proteome changes do not correlate (Jain, 2010; Jain *et al.*, 2011), and thus there may be finer levels of gene regulation at play. Small RNA molecules could be one such mechanism of fine gene regulation. Sequencing of the *C. difficile* 630 genome in 2006 (Sebaihia *et al*.) has made the identification of sRNAs much more straightforward. The first paper published that looked at sRNAs in *C. difficile* was Chen *et al*. in 2011. This study found that the number of sRNAs in the genomes of pathogenic Clostridia is greater than that of non-pathogenic Clostridia. According to this study *C. difficile* contains the greatest number of individual sRNAs of all Clostridia they investigated, 264 in total. The development of the Bacterial Small RNA Database (BSRD) (http://kwanlab.bio.cuhk.edu.hk/BSRD) in 2013 provided a central repository of sRNAs from Gram-positive organisms (both predicted and experimentally verified) and allowed other researchers to design primers for a small number of these sRNAs to study within *C. difficile* (Li *et al*., 2013). Later in 2013, Soutourina *et al.* published their work on sRNAs in *C. difficile* 630, in which they identified over 250 sRNAs by RNAseq; further investigation found growth phase dependent expression of a number of sRNAs in addition to evidence of a regulatory role in motility and biofilm control for 16 c-di-GMP-associated riboswitches (Soutourina *et al.*, 2013). Soutourina *et al.* (2013) also carried out transcription start site (TSS) analysis which we were able to utilise in our RNAseq experiments to identity sRNAs in our samples.

We hypothesise that within the ever-changing gut environment, more precise gene regulation is required and that sRNAs could play a role in pathogenesis, virulence and

stress response (Jain *et al.*, 2011; Ternan, 2013; Ternan *et al*., 2018). There is still much to learn of sRNA activity in *C. difficile*, highlighting the importance of further investigation into how these regulatory RNA factors affect the behaviour of this pathogenic organism. It is especially important to investigate these regulatory RNAs in the presence of faecal water as this allows the replication of a more biologically relevant environment, the type of environment in which *C. difficile* is pathogenic and causes the most harm.

## 5.2 AIMS AND OBJECTIVES

The aim of this chapter is to investigate the expression of sRNAs in *Clostridioides difficile* 630 through the use of RNAseq and RT-qPCR techniques in order to provide a greater insight into how these molecules exert their regulatory abilities within this pathogen, and the wider implications this has on the organism.

The main objectives for this chapter are:

- Design and validate primers for use in RT-qPCR analysis of sRNAs.
- Establish how or if sRNA expression levels differ throughout growth.
- Compare sRNA expression levels in *C. difficile* 630 under normal growth media and in faecal water by RT-qPCR and RNA sequencing.
- Attempt to elucidate links between sRNA and gene expression to gain insight into their function.

## 5.3 RESULTS

### 5.3.1 Growth of *Clostridioides difficile* 630

In order to study the growth of *C. difficile* strain 630 in a more biologically realistic environment, the organism was grown in three separate media i.e. BHIS, BHIS/faecal water (FW) (50/50 v/v), BHIS/PBS (50/50 v/v) as outlined in Chapter 2.2. Attenuance (D650 nm) was measured in triplicate over a time period of seven hours. Statistical analysis showed that the presence of FW in the media was not detrimental to growth of *C. difficile* strain 630 over this period of time (Figure 5.3).

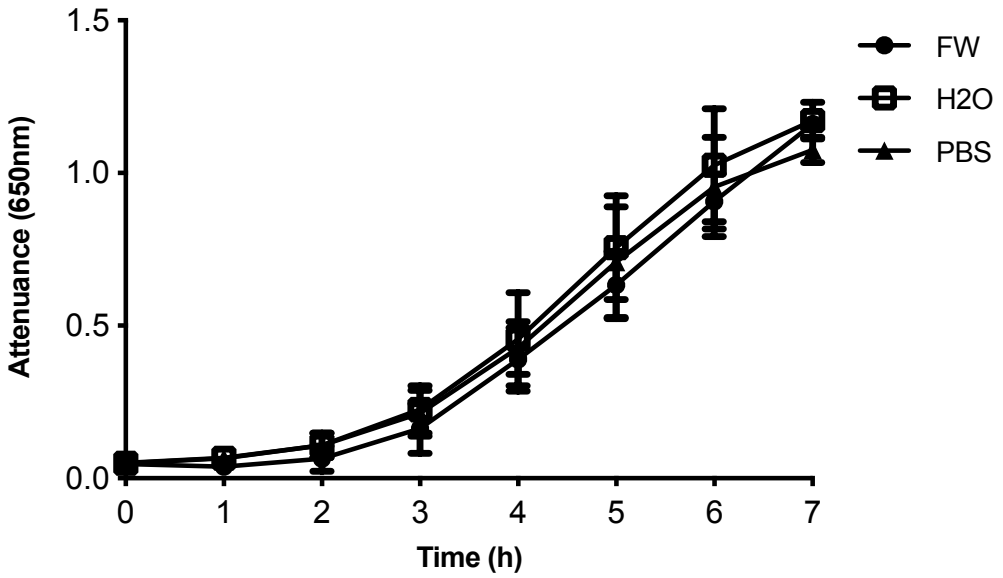



**Figure 5.3:** Growth (D650 nm) of *Clostridioides difficile* strain 630 in BHIS containing 50% faecal water, BHIS containing 50% PBS, and standard BHIS made with deionised $H_2O$. Data presented are means of three independent biological replicates and error bars represent the standard deviation of the mean.

### 5.3.2 sRNA primer validation

Primers were designed for *C. difficile* sRNAs chosen from the Bacterial Small RNA Database (BSRD) (Li *et al.,* 2013), or from Soutourina *et al.* (2013) based on our own RNA sequencing dataset where we saw differential expression in FW (Chapter 3). The sRNAs selected are shown in Table 5.2. It was determined that in order to optimise our chances of accurate amplification, primers would be designed for the ten largest sRNAs in the BSRD. The presence of these was subsequently validated by PCR amplification (Figures 5.4a and b). In addition, six sRNAs were selected from the RNA sequencing data, these were also validated as shown in Figure 5.5.

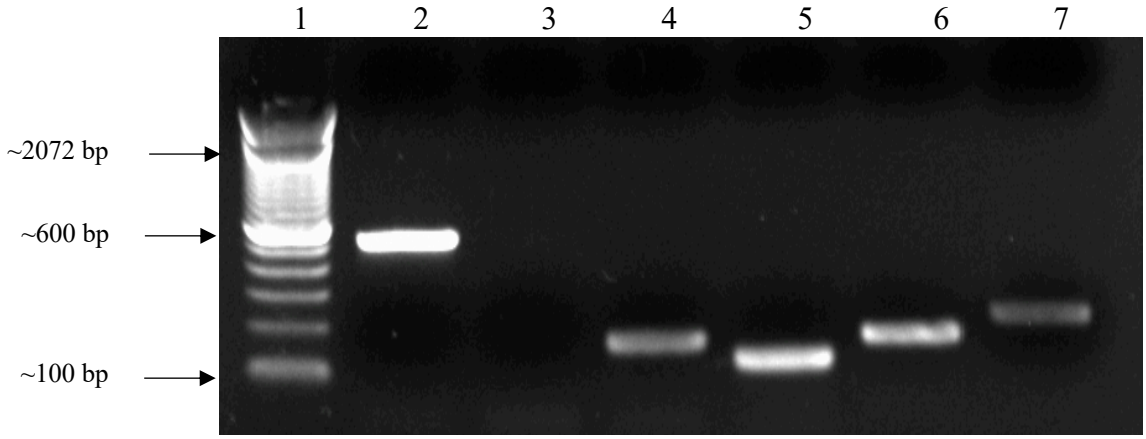



**Figure 5.4a:** PCR validation of sRNA primers with *C. difficile* 630 genomic DNA. **Lane 1:** 100 bp molecular weight marker; **Lane 2:** *tpi* (CD3172); **Lane 3:** *tpi* (CD3172) no template control; **Lane 4:** scdf3688.1; **Lane 5:** scdf3148.1; **Lane 6:** scdf21.1; **Lane 7:** scdf2295.1.

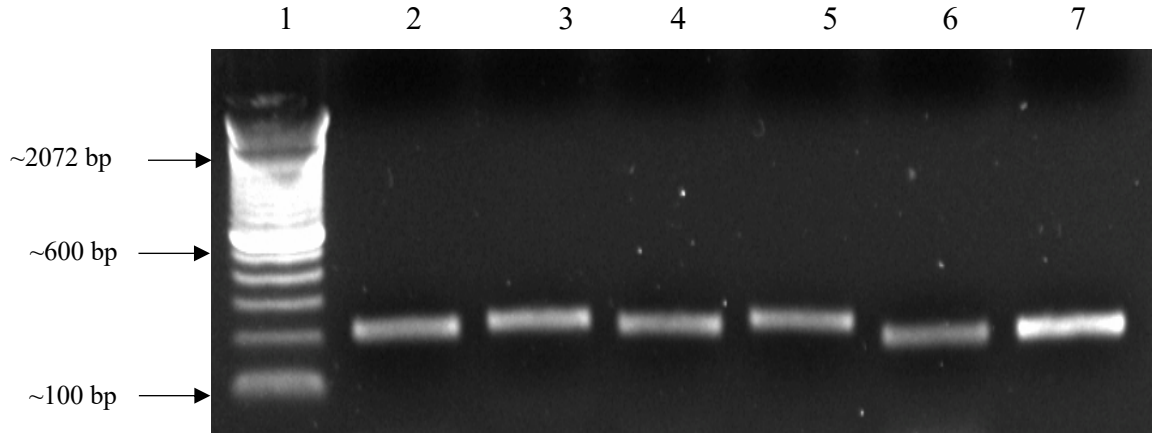



**Figure 5.4b:** PCR validation of sRNA primers with *C. difficile* 630 genomic DNA. **Lane 1:** 100 bp molecular weight marker; **Lane 2:** scdf2056.1; **Lane 3:** scdf3815.1; **Lane 4:** scdf1811.1; **Lane 5:** scdf3115.1; **Lane 6:** scdf1763.1; **Lane 7:** scdf1153.1.

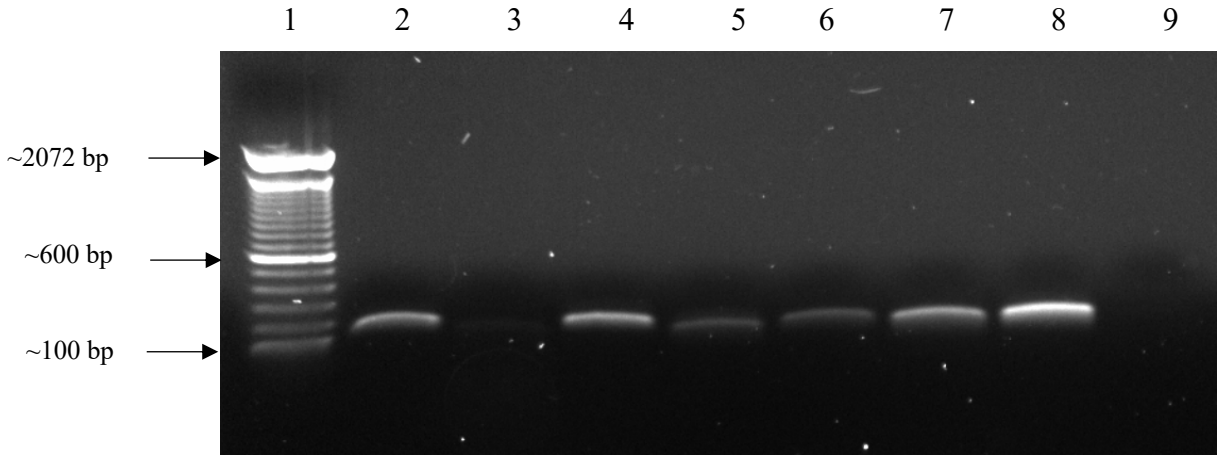



**Figure 5.5:** PCR validation of sRNA primers with *C. difficile* 630 genomic DNA. **Lane 1:** 100 bp molecular weight marker; **Lane 2:** CD630_s0030; **Lane 3:** CD630_n00170; **Lane 4:** CD630_n00330; **Lane 5:** CD630_s0360; **Lane 6:** CD630_s0410; **Lane 7:** CD630_s0490**; Lane 8**: *fliC* positive control; **Lane 9**: *fliC* no template control.

### 5.3.3 Validation of sRNA primers with cDNA

Following reverse transcription of genomic DNA (as outlined in Chapter 3), the sRNA primers were validated by end-point RT-PCR to ensure there was sufficient amplification from the cDNA template (Figures 5.6a and b).

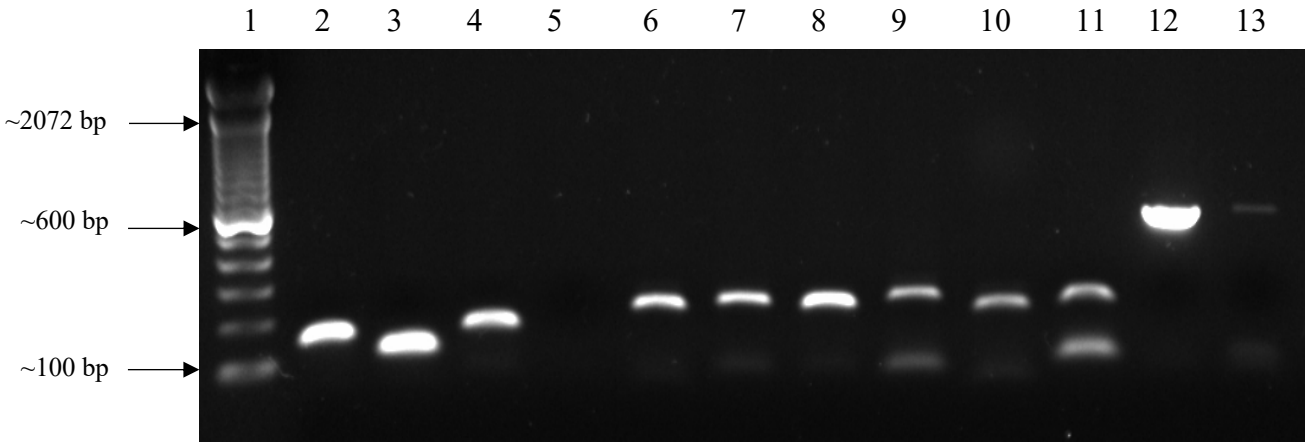



**Figure 5.6a:** RT-PCR of sRNAs from *C. difficile* 630 cDNA. **Lane 1:** 100 bp molecular weight marker; **Lane 2:** scdf3688.1; **Lane 3:** scdf3148.1; **Lane 4:** scdf21.1; **Lane 5:** scdf2295.1 – not amplified (see Figure 5.6b); **Lane 6:** scdf2056.1; **Lane 7**: scdf3815.1; **Lane 8:** scdf1811.1; **Lane 9:** scdf3115.1; **Lane 10:** scdf1763.1; **Lane 11:** scdf1153.1; **Lane 12:** *tpi* (CD3172) gDNA positive PCR control; **Lane 13:** *tpi* (CD3172) no template control.

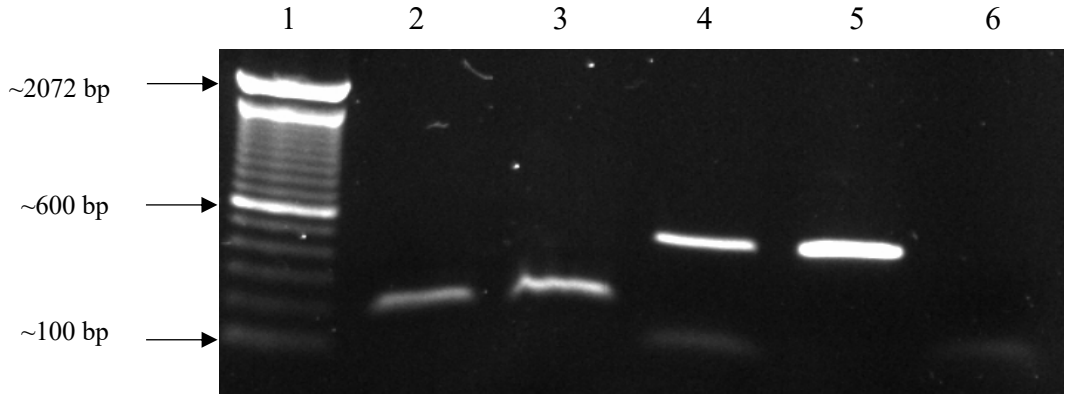



**Figure 5.6b:** RT-PCR of scdf2295.1 from *C. difficile* 630 cDNA. Lane 1: 100 bp molecular weight marker; Lane 2: scdf2295.1; Lane 3: scdf2295.1 gDNA positive control; Lane 4: tpi (CD3172); Lane 5: *tpi* (CD3172) gDNA positive control; Lane 6: *tpi* (CD3172) no template control.

### 5.3.4 sRNA sequencing

As an initial quality control measure, Sanger sequencing was performed using both forward and reverse primers for the sRNAs scdf1811.1 and scdf21.1 with either cDNA, or *C. difficile* 630 genomic DNA as template. Each of the eight samples sequenced matched to the *C. difficile* 630 genome within the range of 97-99% proving that the sRNA sequences had been amplified accurately, and from the correct location within the genome from cDNA and genomic DNA templates. Figure 5.7 is a representative example of an NCBI BLAST alignment for the sequencing product of scdf1811.1 from C. *difficile* 630 cDNA amplified with the reverse primer (Chapter 2, Table 2.3).

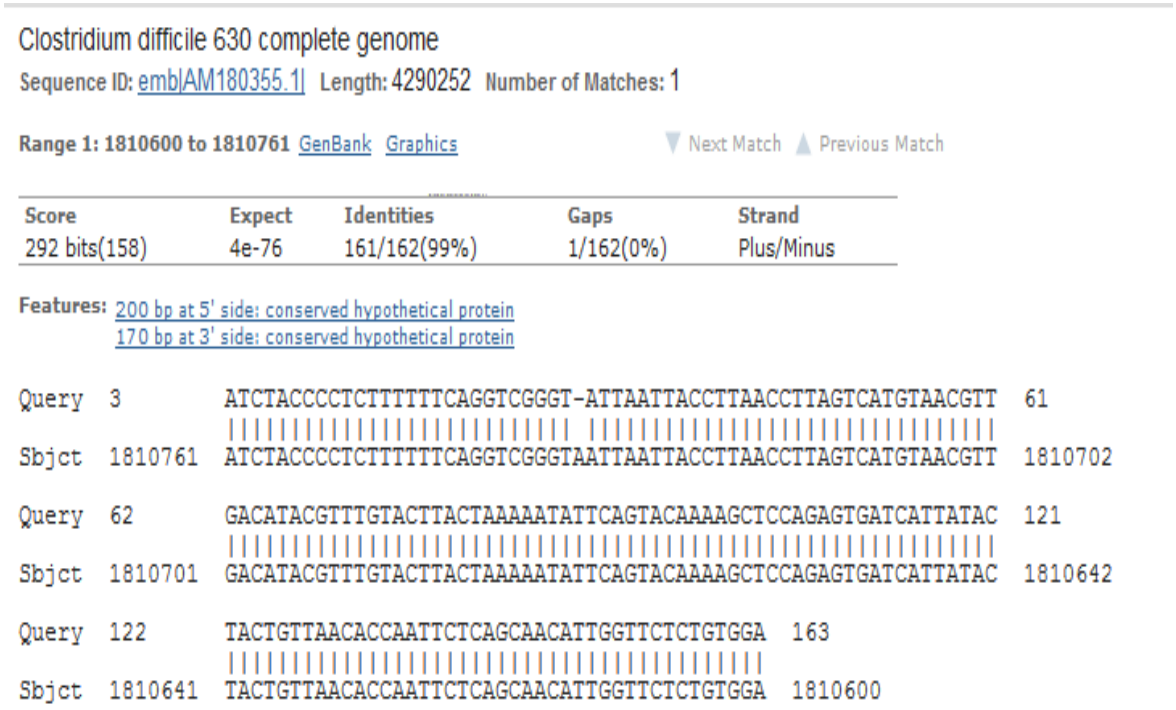



**Figure 5.7:** NCBI sequence alignment of scdf1811.1 amplified from *C. difficile* 630 cDNA (query) against genome reference sequence NC_009089.1.

### 5.3.5 Growth phase dependent expression of sRNAs in *C. difficile*

In an effort to determine whether expression of sRNAs changed throughout growth in *Clostridioides difficile* 630, RT-qPCR was carried out (Figures 5.8a and b). RNA was extracted and reverse transcribed from cell pellets collected at early, mid, late and stationary phases of growth in BHIS broth. The reference gene (Chapter 2, Table 2.1) *rrs* (16S ribosomal RNA) was used as it was shown by Metcalf *et al.* (2010) to be one of the most stable in *Clostridioides difficile* gene expression studies. However, following analysis of several reference genes as outlined in Chapter 3, the three reference genes used in all further RT-qPCR experiments were *adk*, *rpsJ* and *gyrA*. The $\Delta\Delta$ $C_t$ method was used to analyse the data produced from RT-qPCR of scdf1811.1 (Figure 5.8a) and scdf21.1 (Figure 5.8b), this method assumes that the efficiency of the primers is equal to 2 (i.e. that the amount of template is doubled in each cycle). It was decided to further investigate the expression of these two particular sRNAs as they were annotated as having two different functions within the BSRD (scdf1811.1 as a T-box leader and scdf21.1 as a signal recognition particle (SRP). It was found that the expression of scdf1811.1 was more than 3 times higher during mid-log when levels were normalised to early-log phase of growth. In contrast, scdf21.1 expression was lowest during mid-log and more than two and a half times higher during the stationary phase of growth compared to early-log expression levels. These findings show that sRNA expression is changing throughout growth of the organism.

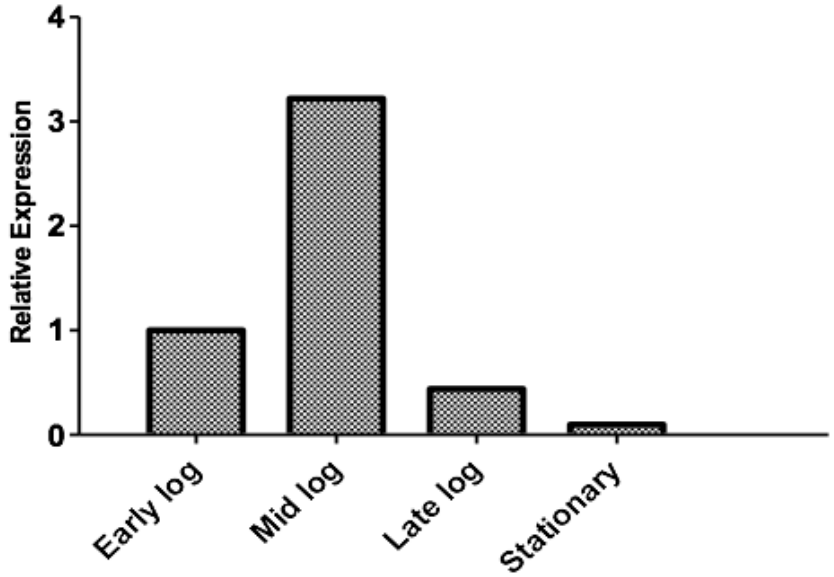



**Figure 5.8a:** Relative quantification analysis of scdf1811.1 expression in *C. difficile* 630 relative to reference gene *rrs* using the ΔΔ $C_t$ method normalised to early-log phase of growth where early-log is set to 1 and the relative expression of all other timepoints are a reflection of this.

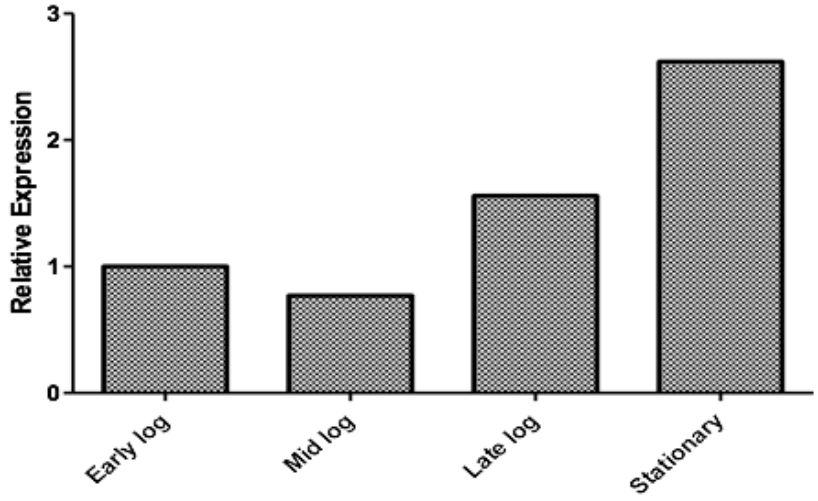



**Figure 5.8b:** Relative quantification analysis of scdf21.1 expression in *C. difficile* 630 relative to reference gene *rrs* using the ΔΔ $C_t$ method normalised to early-log phase of growth where early-log is set to 1 and the relative expression of all other timepoints are a reflection of this.

### 5.3.6 sRNA expression levels in *C. difficile* cells grown in the presence of faecal water

In an effort to understand how sRNA expression changes in the presence of faecal water, *C. difficile* 630 cell pellets were harvested at late-log phase of growth (as previously described). Late-log was chosen as the most appropriate time to collect biomass as this is when the highest abundance of cells can be harvested from cultures grown using the limited supply of faecal water available. It was decided that in order to obtain the most reliable RT-qPCR data, standard curves would be produced for each primer set and RT-qPCR analysis would be conducted using this data (method outlined in Chapter 2). From the standard curve data, the exact efficiency of the primers could then be calculated and should ideally fall between 1.9 and 2.1 (Table 5.1).

**Table 5.1:** Standard curve data for all genes studied by RT-qPCR in this chapter.

| | **Error** | **Efficiency** | **Slope** | **Y-intercept** |
|---|---|---|---|---|
| ***scdf3688.1*** | 0.046 | 1.974 | -3.386 | 9.88 |
| ***scdf3148.1*** | 0.041 | 1.872 | -3.671 | 10.48 |
| ***scdf21.1*** | 0.033 | 1.963 | -3.414 | 19.13 |
| ***scdf2295.1*** | 0.015 | 2.036 | -3.240 | 20.71 |
| ***scdf2056.1*** | 0.041 | 2.066 | -3.173 | 23.34 |
| ***scdf1811.1*** | 0.067 | 2.020 | -3.275 | 24.55 |
| ***scdf3115.1*** | 0.007 | 2.104 | -3.095 | 21.39 |
| ***CD630_s0490*** | 0.034 | 1.926 | -3.514 | 24.61 |
| ***CD630_s0030*** | 0.089 | 2.030 | -3.252 | 20.09 |
| ***CD630_n00330*** | 0.016 | 1.906 | -3.569 | 26.92 |
| ***CD630_s0360*** | 0.024 | 1.944 | -3.464 | 25.56 |

The relative expression levels of the target genes i.e. sRNAs can then be determined from the known standards. From the initial ten sRNAs, suitable standard curves were produced for eight of them (Table 5.1). All data was normalised to the BHIS control and a combination of the reference genes *adk*, *rpsJ* and *gyrA* (see standard curve data in Chapter 3, Table 3.3) was used to calculate the fold change. The analysis showed that expression of three of the sRNAs scdf3688.1, scdf3115.1 and scdf1763.1 were 1.36, 2.12 and 1.25-fold down, respectively, in faecal water. Expression of five of the

sRNAs was increased as follows scdf3148.1 by 1.19, scdf21.1 by 1.4, scdf2295.1 by 2.23, scdf2056.1 by 3.51 and scdf1811.1 by 1.48-fold (Figure 5.9).

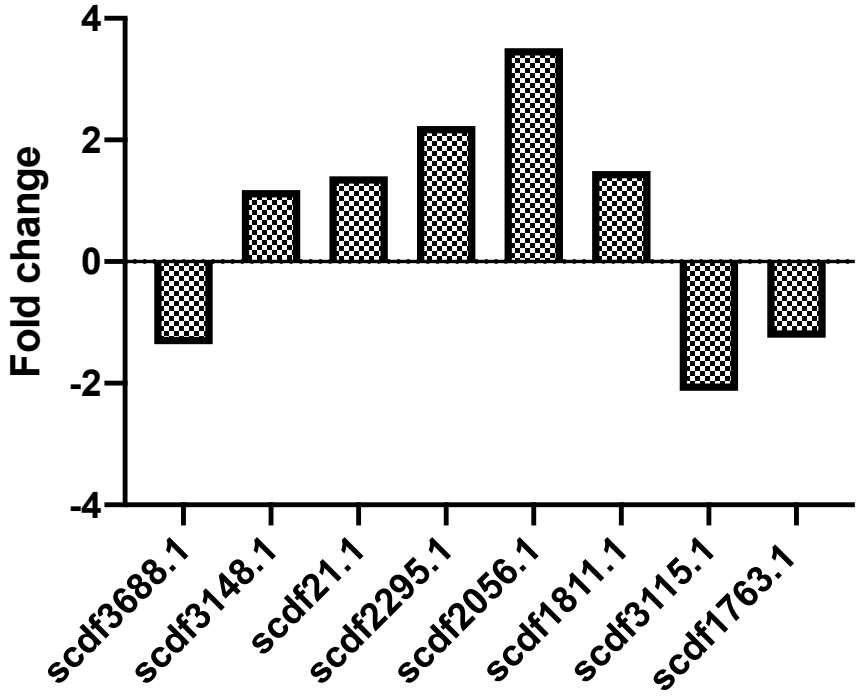



**Figure 5.9:** Fold-change of sRNAs from RT-qPCR of *C. difficile* 630 grown in faecal water and BHIS control collected at late-log phase of growth. Data is normalised to the reference genes *adk*, *rpsJ* and *gyrA*.

### 5.3.7 Verification of RNA sequencing results with RT-qPCR

From the RNA sequencing data obtained (Chapter 3) it was noted that 51 sRNAs (p-value < 0.01 and fold-change > 1.43) were differentially expressed in the faecal water grown samples. In order to further verify these results a number of these sRNAs were chosen for RT-qPCR analysis. We found that direction of expression i.e. up or down, was the same in both datasets for the five sRNAs studied (Figure 5.10). This helped to

further reinforce our confidence in the RNA sequencing and RT-qPCR data obtained to this point.

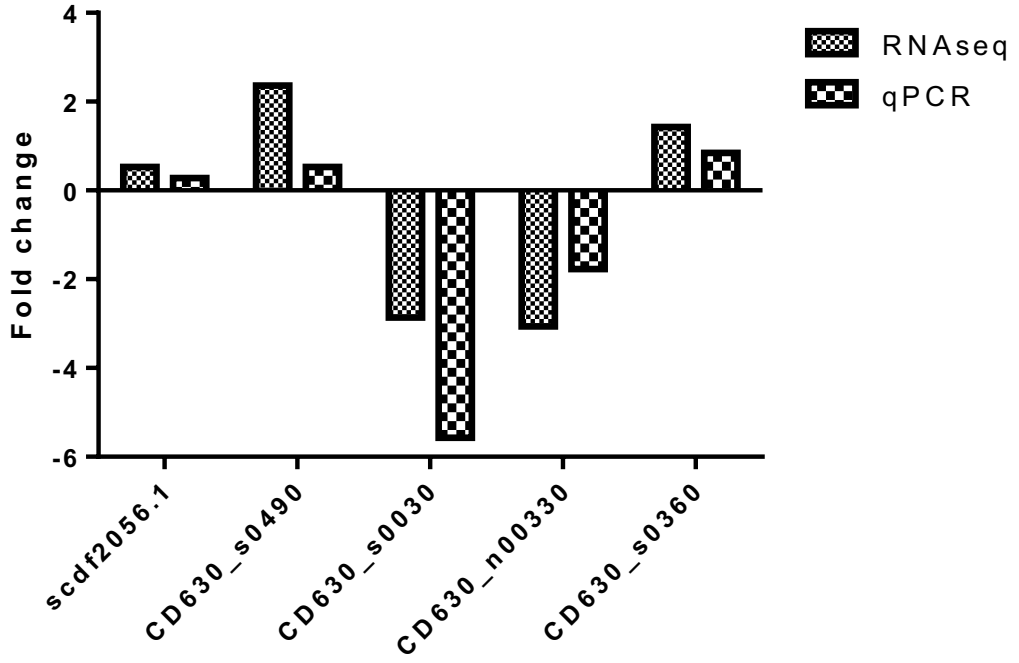



**Figure 5.10:** Fold change of sRNAs from *C. difficile* 630 grown in faecal water from RNA sequencing data compared against the fold change obtained from RT-qPCR. RT-qPCR data was normalised to the reference genes *adk*, *rpsJ* and *gyrA*.

### 5.3.8 Antisense sRNA and corresponding gene analysis

The RNAseq dataset was further analysed by selecting the sRNAs that were annotated as antisense by Soutourina *et al*. (2013) thus allowing a specific gene to be assigned to each sRNA as the target of action of the sRNA i.e. the gene complementary to the antisense sRNA. Antisense sRNAs are located on the "antisense" strand opposite the "sense" coding strand either within the coding sequence or the 5' or 3' UTRs. They are fully complementary to the coding mRNA and that's how they are able to bind and

exert their effect - the majority of antisense RNAs are cis-acting for this reason. IGR located sRNAs tend to be trans-acting and require Hfq for example, we haven't gone into these here as they are beyond the scope of this chapter. Small RNAs were assigned to genes using the start and end coordinates of each sRNA to search the *C. difficile* 630 NCBI reference sequence NC_009089.1. From this, 9 antisense RNAs were matched to their complementary genes (Table 5.3/Figure 5.11). Of these 9, 4 showed decreased expression in faecal water and 5 showed increased expression. Three of the sRNAs had corresponding genes that had p-values of 0.09 or greater (and therefore outside of the cut-off value of < 0.01), therefore these 3 pairs were eliminated from further study. The remaining 6 sRNA/gene pairs were expressed as follows: CD630_n00080 and its corresponding gene *fliG* were both down by 3.64 and 1.67-fold, respectively. CD630_n00090 and *fliH* were down by 2.37 and 1.65-fold. CD630_n00450 and *norV* were down by 2.16 and 2.18-fold. SQ2179 was down by 3.55-fold, whereas its corresponding gene *xylR* was up by 1.68-fold. CD630_n01110 and *SpoIIE* were both increased by 4.47 and 46.14-fold, respectively. CD630_n00650 and CD1893 were also both up in the faecal water samples by 2.91 and 4.17-fold.

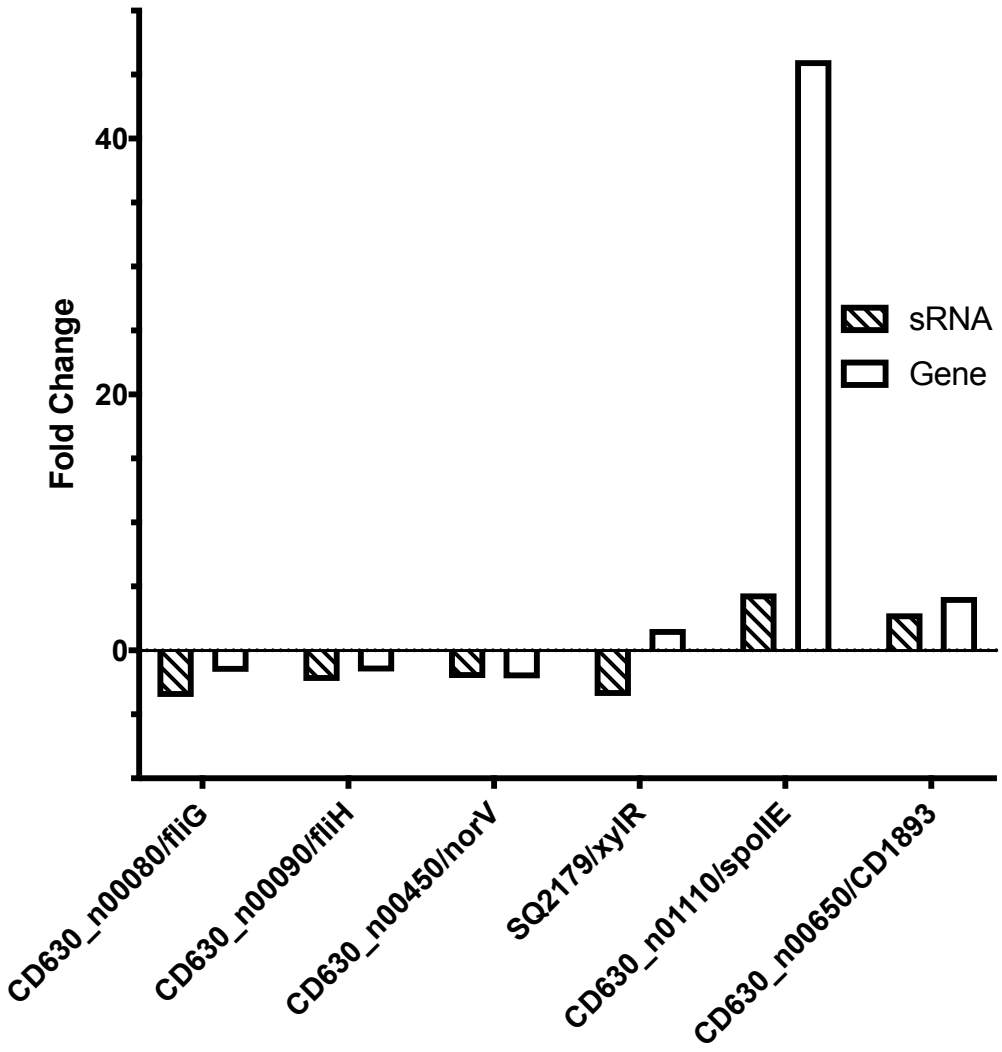



**Figure 5.11:** RNA sequencing fold change of sRNA and corresponding gene from *C. difficile* 630 grown in faecal water. All sRNAs are antisense to their corresponding gene. Only sRNAs/genes with p-value < 0.01 from RNAseq data included.

**Table 5.2:** sRNAs from Bacterial Small RNA Database (BSRD) including related information from Soutourina *et al*., 2013 where available.

| sRNA ID | Function | Length | Start | End | Strand | Soutourina sRNA ID | Soutourina Function (Rfam) | Soutourina Length | Start | End |
|---|---|---|---|---|---|---|---|---|---|---|
| **scdf3688.1** | SsrA | 350 | 3687389 | 3687738 | -1 | CD630_s0600 | tmRNA | 350 | 3687389 | 3687738 |
| **scdf3148.1** | RNaseP_bact_a | 335 | 3147911 | 3148245 | -1 | CD630_s0570 | Rnase P | 211 | 3148001 | 3148211 |
| **scdf21.1** | ffs | 265 | 20129 | 20393 | 1 | No match | | | | |
| **scdf2295.1** | T-box | 264 | 2294414 | 2294677 | -1 | CD630_s0440 | T box (Trp) | 279 | 2294385 | 2294663 |
| **scdf2056.1** | T-box | 263 | 2055449 | 2055711 | 1 | CD630_s0390 | T box | 279 | 2055462 | 2055740 |
| **scdf3815.1** | T-box | 263 | 3814001 | 3814263 | -1 | | | | | |
| **scdf1811.1** | T-box | 258 | 1810561 | 1810818 | 1 | SQ1043 | | | | |
| **scdf3115.1** | T-box | 256 | 3114822 | 3115077 | -1 | CD630_s0560 | T box (Asn) | 273 | 3114793 | 3115065 |
| **scdf1763.1** | T-box | 253 | 1762876 | 1763128 | 1 | SQ1005 | Putative sRNA candidate | 243 | 1762914 | 1763156 |
| **scdf1153.1** | T-box | 247 | 1152611 | 1152857 | 1 | | | | | |

**Table 5.3:** sRNAs from RNAseq data whose expression was altered in response to faecal water (FW), and their corresponding gene, p-value < 0.01.

| id | Fold Change | Expression change in FW | Start | End | TSS | Strand | Annotation | Corresponding gene | Corresponding gene function | Corresponding gene in DE data |
|---|---|---|---|---|---|---|---|---|---|---|
| **CD630_n00080** | 3.64 | down | 312822 | 312924 | 312924 | -1 | Antisense CDS | *fliG* | flagellar motor switch protein | 1.67 down |
| **CD630_n00090** | 2.37 | down | 313159 | 313261 | 313261 | -1 | Antisense CDS | *fliH* | flagellar assembly protein | 1.65 down |
| **CD630_n00450** | 2.16 | down | 1355629 | 1355729 | 1355729 | -1 | Antisense CDS | *norV* | anaerobic nitric oxide reductase flavorubredoxin | 2.18 down |
| **SQ2179** | 3.55 | down | 3562943 | 3563188 | 3563322 | -1 | Antisense 3'UTR | *xylR* | xylose repressor | 1.68 up |
| **CD630_n01110** | 4.47 | up | 4081910 | 4082308 | 4081910 | 1 | Antisense CDS | *spoIIE* | stage II sporulation protein | 46.14 up |
| **CD630_n00650** | 2.91 | up | 2198177 | 2198277 | 2198277 | -1 | Antisense CDS | CD1893 | oligonucleotide binding regulator | 4.17 up |

## 5.4 DISCUSSION

There is an abundance of literature available which confirms that small RNAs (sRNAs) are found in all kingdoms of life and are now acknowledged as a unique class of gene expression modulators. Recent efforts have focused on determining sRNA functions under various physiological conditions and the techniques they use to exert these effects (Michaux *et al*., 2014). One such class of sRNA is the riboswitch, which acts by sensing ions, metabolites, temperature and pH changes and then assuming alternate configurations in response to these, in order to regulate downstream activities such as RNA processing and stability, transcription termination and translation initiation (Sedlyarova *et al*., 2016; Mellin and Cossart, 2015). This chapter contributes to the current understanding of how *cis*-acting sRNAs, which are mostly found in the 5' untranslated region (UTR), 3' UTR or coding sequence (CDS) antisense to their target gene, behave in *Clostridioides difficile* 630 and potentially contribute to pathogenesis in this organism.

### 5.4.1 sRNA expression is growth phase dependent

The small RNA scdf1811.1 in *C. difficile* 630 is a riboswitch type sRNA which acts as a T-box leader according to the BSRD (Li *et al.*, 2013). T-box leaders are regulatory RNA molecules which are commonly used by Gram-positive organisms to control the expression of aminoacyl-tRNA synthetase genes as well as those associated with amino acid biosynthesis and uptake (Green *et al.*, 2010). They work by recognising specific tRNAs and undergoing a conformational change upon binding, this leads to either initiation or inhibition at the level of transcription or translation of the downstream gene (Stamatopoulou *et al.*, 2017). It was found that the expression of

scdf1811.1 was more than 3 times higher during mid-log phase of growth when levels were corrected to early-log phase. It could be suggested that this sRNA is highly expressed during mid-log as it may be involved in the expression of genes involved in the exponential growth phase of *C. difficile* 630 such as those involved in amino acid biosynthesis - aminoacyl-tRNA synthetases for example. Extensive research has been conducted in the model organism *Bacillus subtilis* into how T-box riboswitches are involved in gene regulation. One particular gene that has been studied is the *tyrS* gene, which is an aminoacyl-tRNA synthetase gene and encodes tyrosyl-tRNA synthetase (TyrTS); it has been found that this gene (and others like it) are induced by limitation of the specific amino acid and regulated by transcription antitermination. Transcription antitermination is when the RNA polymerase continues elongation of the transcript by ignoring the termination signal allowing transcription of the section of RNA beyond the terminator (Henkin *et al.*, 1992; Putzer *et al.*, 1992; Grundy *et al.*, 1994). T-box riboswitches are able to sense which amino acids are present for the ribosome to translate by analysing the aminoacylation levels of tRNAs. The T box leader's RNA secondary structure consists of three stem loops (I, II and III) and a pseudoknot, which bind to the target tRNA and subsequently undergo conformational change to either prevent or initiate transcription termination (Figure 5.2) (Zhang and Ferré-D'Amaré, 2015). In the case of our data it could be that transcription antitermination is allowing the transcription of amino-acid biosynthesis related genes during exponential growth due to increased demand. Regulation of tRNA synthetase genes has been studied in *E. coli* and it was found that expression is coupled to the growth rate of the cell (Grunberg-Manago *et al.*, 1996). It could be suggested from the data in this study that as amino acids are in high demand due to the level of protein synthesis required during this period of rapid growth in *C. difficile* 630, tight regulation is imperative to ensure

accurate and efficient replication, which is where this particular T-box leader riboswitch comes in. The data presented here goes towards the understanding that riboswitches play a significant role in the gene expression of *C. difficile* 630 and in particular during exponential growth, when a large amount of gene regulation is likely to be required.

In contrast, the expression of scdf21.1, which is annotated as a signal recognition particle (SRP) in the BSRD (Li *et al.*, 2013), was highest during the stationary-phase of growth, with levels more than two and a half times early-log expression. SRPs are universally conserved ribonucleoproteins (protein-RNA complexes) that in prokaryotes recognise and bind specific proteins in order to guide them to the plasma membrane to ensure they find their proper cellular location (Akopian *et al.*, 2013). The SRP is thought to be composed of two main parts, the fifty-four homologue (Ffh) protein and 4.5S RNA (the small cytoplasmic RNA, or scRNA) (Patenge *et al.*, 2012). A recent study described how the 4.5S RNA, a component of the SRP, has an influence on virulence in the Gram-positive human pathogen *Streptococcus pyogenes* (cause of Group A strep). It was found that mutation of the 4.5S RNA gene negatively impacted bacterial growth, reduced virulence factor expression, and lowered virulence in a mouse infection model (Trevino *et al.*, 2010). Rosch *et al.* (2008) reported that a number of virulence factors involved in host cell-pathogen interactions required the SRP pathway for secretion in *Streptococcus pyogenes*. It could be suggested that the increase in expression levels during stationary phase of growth in *C. difficile* 630 is associated with virulence gene expression. For example, the toxin genes *tcdA*, *tcdB* and *tcdR* are transcribed during the stationary growth phase (Dupuy *et al*., 2008)

therefore it could be hypothesised that this particular sRNA plays a role in this type of gene regulation. The SRP pathway also plays a role in spore formation in *B.* subtilis; Yamane *et al*. (2004) found that impairment of the SRP complex pathway, by construction of a FtsY (SRP membrane receptor filament) mutant, resulted in a 70% reduction in spore formation. It is worth considering the possibility that an increase in this particular SRP sRNA in *C. difficile* 630 during stationary phase of growth is involved in the sporulation process which would be beginning to take place at this stage of growth. Further study into the relationship between SRPs and virulence in *C. difficile* 630 would be required to elucidate a potential link between the two. Soutourina *et al.* (2013) have also shown that a number of sRNAs are growth-phase dependent in *C. difficile*. They found three sRNAs with their highest expression levels during exponential growth and three highest during stationary phase. The work presented here mirrors the findings of Soutourina *et al*. showing that sRNAs are differentially expressed throughout growth in *C. difficile* 630.

### 5.4.2 sRNA expression changes when exposed to faecal water enriched growth medium

Preliminary investigations by RT-qPCR indicated that sRNA levels change when *C. difficile* 630 is grown in faecal water enriched growth medium. Previous results have shown that sporulation is increased in *C. difficile* 630 in the presence of faecal water, therefore, it is reasonable to suggest that the organism senses a change in environment and responds accordingly. There have been numerous studies in recent years that have investigated the important role of sRNAs in environmental response and survival in bacterial species (Altuvia *et al.*, 1997; Majdalani *et al.*, 1998; Delihas and Forst, 2001;

Majdalani *et al.*, 2001; Masse and Gottesman, 2002; Lease *et al.*, 2004; Bak *et al.*, 2014). From eight sRNAs initially studied, three were found to exhibit reduced expression in the presence of faecal water. It could be hypothesised that the downregulation of some sRNAs in the presence of faecal water allows the target gene to be expressed undisturbed, as many sRNAs are shown to act as inhibitors of translation by base pairing with the mRNA in the ribosome-binding site (Gottesman, 2005). The remaining five sRNAs showed increased expression in the faecal water grown samples. A study by Michaux *et al.* in 2014 found that virulence was increased in *Enterococcus faecalis* when sRNA deletion mutant strains were used to infect larvae and mouse models; an sRNA homologous to the RNAII component of the toxin-antitoxin system (induced by stress) in particular was found be heavily involved in virulence within this organism. In addition to this, a study by Rau *et al*. (2015) confirmed that sRNAs were involved in regulatory processes within *E. coli* and found that the majority of sRNAs upregulated under multiple chemical stress conditions were involved in the repression of outer membrane protein expression, possibly due to the organism's physiological response seeking to reduce the entry of toxic extracellular molecules.

Further investigation with the use of RNA sequencing confirmed that sRNA expression levels do change in the presence of faecal water. These results were verified with RT-qPCR to confirm the sequencing results, this helped to further enhance confidence that these changes were happening. By concentrating on a number of antisense sRNAs we were able to analyse the corresponding gene; this helped to assign a potential function to the sRNA. The current literature states that antisense sRNAs

tend to act on the gene that is located on the complementary strand (Thomason and Storz, 2010), therefore we were able to assess the expression levels of the complementary genes from the seq data and determine how/if the sRNA was having a direct effect. The sRNA CD630_n00080 and its corresponding gene *fliG* (encoding a flagellar motor switch protein) were both down by 3.64 and 1.67-fold, respectively. In addition to this the sRNA CD630_n00090 and the complementary gene *fliH* (encoding a flagellar assembly protein) were both down by 2.37 and 1.65-fold. The data shows that for both pairs the sRNA exhibits decreased expression by similar amounts to the gene, this may suggest a close link between the expression of the two RNAs, and that the corresponding sRNA normally plays a role in the expression of these two motility genes. As flagellar gene expression occurs in highly controlled stages (Pallen and Matzke, 2006) it could be said that the sRNA is not yet needed at this particular time and has therefore not yet been switched on. We know from our earlier work (Chapter 3; Ternan *et al*., 2018) that motility gene expression is reduced in *C. difficile* 630 grown in the presence of faecal water, thereby it is fair to suggest that sRNAs may play a role in this. Previous work on *E. coli* has uncovered a link between sRNAs and motility gene expression (De Lay and Gottesman, 2012; Bak *et al*., 2015). sRNAs have been shown to both positively and negatively regulate the main genes responsible for flagellar synthesis in *E. coli*, *flhD* and *flhC* (De Lay and Gottesman, 2012). Soutourina *et al.* (2013) first highlighted the role of sRNAs in *C. difficile* motility regulation on discovery of a group of sRNAs that respond to the universal signaling molecule c-di-GMP (cyclic-di-guanosyl-5'monophosphate) and that these particular sRNAs affected motility as well and biofilm formation and virulence.

The gene *norV*, a putative anaerobic nitric oxide reductase flavorubredoxin is complementary to the antisense sRNA CD630_n00450, they were both decreased by 2.18 and 2.16-fold in the faecal water samples. NorV has been shown to be involved in the metabolism of nitric oxide (NO) in *E. coli* cells grown anaerobically (Hutchings *et al*., 2002). Nitric oxide (NO) is a key signalling molecule in the gastrointestinal tract and has been shown to be involved in blood flow, smooth muscle contraction and relaxation as well as secretory and immunological control (Vermeiren *et al*., 2009). Research has shown that bilirubin acts as an endogenous scavenger of NO within the gut thereby providing a protective mechanism to the presence of reactive species (Mancuso *et al*., 2006). It could therefore be suggested that NorV as a NO detoxifying enzyme may be downregulated in the presence of bilirubin (within the faecal water milieu) as NO detoxification is no longer required by *C. difficile* in this particular growth environment (Kochar *et al*., 2011).

The sRNA SQ2179 was decreased by 3.55-fold whereas its corresponding gene *xylR* encoding a xylose repressor was increased by 1.68-fold. Xylose is a monosaccharide used as a carbon and energy source by many microorganisms (Heo *et al*., 2008). XylR has been shown to positively regulate xyl genes in *E. coli* and *Salmonella typhimurium* and negatively regulate xyl genes in other organisms including *Bacillus* and *Lactobacillus* species (Heo *et al*., 2008). The reduction in expression of the sRNA may have led to an increase in expression of the *xylR* gene as the sRNA's regulatory effect was absent. As XylR is itself involved in the regulation of xylose metabolism, it would suggest that this pathway is affected by the presence of faecal water. It stands to reason

that these regulatory effects may be due to an adaptive response of the organism to other carbon sources and additional dietary metabolites present in the faecal water.

CD630_n01110 and *SpoIIE* (encodes stage II sporulation protein e) were both increased by 4.47 and 46.14-fold, respectively. We have shown previously (Chapter 3; Ternan *et al.*, 2018) that sporulation is massively upregulated in *C. difficile* 630 grown in faecal water. We therefore hypothesis that a small regulatory RNA may play a role in fine tuning this increase in some way, potentially through transcription antitermination for example. These results show that both *SpoIIE* and the sRNA show increased expression, there is however a 10-fold difference between the two. It is well known that *Bacillus* and *Clostridium* organisms begin sporulating when potentially hostile growth conditions are sensed. The sporulation process is a tightly controlled sequence of events at both the transcriptional and posttranslational levels involving a plethora of additional regulatory factors including sigma, transcription factors as well as many other proteins (Al-Hinai *et al*., 2015). Therefore, it is reasonable to assume that *C. difficile* 630 detects the faecal water environment as it would do in the gut of a vulnerable patient and responds by inducing sporulation. In addition to this, the antisense sRNA CD630_n00650 found to be complementary to the gene CD1893 were both increased in the faecal water samples by 2.91 and 4.17-fold respectively. CD1893 is a putative oligonucleotide binding regulator and has been identified as being regulated by Spo0A (Pettit *et al*., 2014) further highlighting a potential involvement of sRNAs in sporulation in *C. difficile* 630. Small RNA molecules have been found to be involved in sporulation previously such as the sRNA CsfG, which has been found to be required for effective spore germination in *B. subtillis* by Marchais *et al*.; this

particular small RNA is transcribed only during sporulation and specifically in the forespore (Marchais *et al*., 2011).

This is the first investigation to show that expression levels of sRNA molecules change during growth in a biologically relevant growth medium i.e. faecal water and that these changes may in turn have significant effects on the expression of other genes potentially involved in virulence of the organism such as sporulation, motility etc. In the current era of antibiotic resistance it has never been more important to uncover potential therapeutic and diagnostic targets such as small RNAs. This work could be further expanded upon with the use of Northern blotting for confirmation of RNA expression and Western blotting to confirm protein abundance and also the effect of sRNAs on the translation of proteins from genes of interest. In addition to this, the development of sRNA mutants would provide a much deeper insight into the effects of these regulatory molecules on *C. difficile* 630 as a whole.

# CHAPTER 6: FINAL DISCUSSION AND FUTURE PERSPECTIVES

*Clostridioides difficile* Infection (CDI) is a known healthcare acquired disease and significant source of antibiotic-associated diarrhoea (AAD) and colitis (Bartlett, 2009). CDI is reported to be responsible for up to a third of recorded instances of AAD worldwide (Arriola *et al*., 2016). The financial burden of CDI on the UK health service is substantial, with a median total cost per patient of between £5631 and £8542 (Wilcox *et al*., 2017). Most of the research conducted into how *C. difficile* causes disease is carried out in artificial laboratory media, limiting insight into how this pathogen responds to the environment in which it causes disease i.e. the large intestine. The overall aim of this study was to construct a faecal water model to study the transcriptome of *C. difficile* 630 in order to characterise the molecular mechanisms underpinning pathogenesis within the organism. We hypothesised that understanding how *C. difficile* functions in a more biologically relevant setting would uncover mechanisms at work that could not be determined with the use of normal laboratory media by developing an *in-vitro* model to determine growth characteristics and the transcriptional profile of *C. difficile* 630 and the mutant strains *Δerm*, *dnaK* and *spo0A*.

Chapter 3 aimed to determine which genes were significantly differentially expressed in the faecal water media in order to characterise the effect of FW on *C. difficile* 630 gene expression by RNAseq and RT-qPCR analysis. We therefore extracted high quality RNA in sufficient quantities to guarantee the reliability of downstream results. Optimisation of the RNA extraction protocol alongside reference gene analysis and adherence to the MIQE guidelines ensured that the transcriptomic data obtained was of optimum quality. The MIQE guidelines are a set of recommendations regarding

RNA quality and RT-qPCR experiment construction and analysis that aim to ensure accuracy and reliability of downstream RT-qPCR results.

The work of Browne *et al*. (2016) showed that the human gut microbiota has extensive sporulation capability, this allowed us to hypothesis that sporulation associated genes would be upregulated in the FW media. The RNAseq results obtained confirmed this, with up-to 300-fold increases strongly indicating that FW is an particularly potent inducer of sporulation in *C. difficile* 630. Increased cell length was observed which we assume may be a prelude to sporulation. Significant modulation of transport system associated genes was observed in the FW samples which could be the organisms attempt to adapt to the increase in carbon sources and other diet-derived metabolites within the FW media. The lack of classical heat-stress response also confirmed that the changes observed are most likely due to the presence of FW and its constituents. The data also revealed an increase in expression of the lantibiotic / multidrug ABC transporters, previous research has shown that the proteins cprA and cprB are involved in resistance to cationic antimicrobial peptides (CAMPs) (McBride and Sonenshein, 2011). We also found an overall reduction in motility gene expression as the flagellar operon was decreased in response to FW. Dingle *et al*. (2011) found that a *C. difficile fliC* mutant strain showed improved adherence to intestinal-derived Caco-2 cells suggesting that flagella are either not needed for virulence or that reduction of motility is a tool utilised to increase pathogenesis of the organism. It could also be suggested that suppression of motility occurs as a form of energy conservation in the gut environment in response to the presence of a semi-solid substance.

Our observations from the RNAseq data led us to explore the expression of key genes associated with motility, sporulation, toxin production and chaperone activity in the two *C. difficile* mutant strains *dnaK* and *spo0A* compared to the wildtype *C. difficile* 630 and the parental strain *Δerm*. We sought to determine if faecal water had an effect on the growth of *Δerm* and the *dnaK* and *spo0A* mutant strains compared to what we witnessed with *C. difficile* 630 in Chapter 3. We found that we were able to replicate the findings of Chapter 3 with 630 as it again grew unaffected by the presence of FW. The mutant strains however, displayed reduced growth rate in the FW compared to the parental *Δerm* strain. Reduced growth rate in *dnaK* mutants has been witnessed previously (Jain *et al.*, 2017) and has been attributed to a deficiency in the cellular synthesis of DNA and RNA (observed by Itikawa and Ryu (1979) in *E. coli dnaK* mutants). Selby *et al*. (2011) found reduced growth rates and lowered stress tolerance in a *C. botulinum dnaK* ClosTron mutant. Our findings indicate that the *dnaK* mutant may perceive the FW environment as stressful. In addition to the observations discussed the *dnaK* mutant also exhibits temperature sensitivity and increased expression of other chaperones (Jain *et al*., 2017) suggesting that it is, physiologically, in a form of permanent heat stress, thereby, making it more sensitive to additional stresses such as the FW environment.

Reduced growth was also seen in the *spo0A* mutant grown in FW, highlighting a potential role for Spo0A in how the organism responds to the intestinal environment. An increase in cell length was found in both the *dnaK* and *spo0A* mutant strains in FW, which we hypothesise could be a possible defence mechanism employed by the organism in response to FW components. The molecular chaperones *groEL* and *groES*

were found to be increased in the *Δerm* and *dnaK* mutant strains and reduced in 630 and *spo0A* strains in FW. As GroEL is an important component of the natural stress response within prokaryotes (Hennequin *et al*., 2001a; Arora *et al*., 2017) our findings suggest that FW appears to have induced a stress-response within the *dnaK* mutant organism. We know that Spo0A regulates processes other than sporulation, such as metabolism and efflux pumps (Molle *et al*., 2003; Deakin *et al*., 2012), it could therefore be said that this increase in GroEL and GroES in the *spo0A* mutant suggests a potential regulatory role for Spo0A in chaperone expression. The expression of *tcdA* was reduced in both the *dnaK* and *spo0A* mutant strains grown in FW. Previous research has linked toxin expression to sporulation, where toxin production is shown to be reduced (Underwood *et al*., 2009) or increased (Deakin *et al*., 2012; Pettit *et al*., 2014) in *spo0A* mutant strains. Other research has found that toxin production can be either increased or decreased in *spo0A* mutants depending on strain type (Mackin *et al*., 2013). Our results add to this knowledge and further highlight the influence of FW on virulence gene expression in this organism. These observations further emphasise the complex nature of *C. difficile* and suggest that there are many elements involved in the expression of these important virulence factors. Differences in sporulation associated gene expression within the *dnaK* mutant strain were observed with *SpoVB* and *cspC* reduced by up to 3-fold. We hypothesis that the *dnaK* mutant reaches stage 5 of sporulation at a later stage than the wild-type and parental strains, this is in keeping with the reduced growth kinetics observed in this strain. This differential expression of the sporulation-associated genes indicates a potential regulatory function of the *dnaK* chaperone with regards to sporulation in the presence of FW within *C. difficile*. This work represents the first study into how FW affects the *C. difficile* mutant strains *dnaK* and s*po0A* and highlights clear differences in the behaviour of the mutant strains

strongly indicating not only that FW has an effect on *C. difficile* and thereby virulance and pathogenesis promoting capabilities but that the genes *spo0A* and *dnaK* may play important roles in the regulation of these responses.

Chapter 5 sought to investigate the expression of sRNAs in *C. difficile* 630 grown in FW through the use of RNAseq and RT-qPCR in order to provide a greater insight into how these molecules exert their regulatory functions within this pathogen, and the wider implications this has for the organism. Initial investigations found that sRNA expression differed throughout growth with expression of a T-box leader riboswitch 3 times higher during mid-log compared to early-log phase. We suggest this increase during exponential growth may be due to the increased demand for amino acid biosynthesis during this phase of growth as previous work by Grunberg-Manago *et al.* (1996) found that expression of tRNA synthetase genes is linked to the growth rate of the cell in *E. coli*. We also found that a sRNA annotated as a signal recognition particle (SRP) was most highly expressed during stationary growth. A number of studies have found a link between SRPs and virulence in bacteria (Yamane *et al*., 2004; Rosch *et al.*, 2008; Trevino *et al.*, 2010); as spore production is active during stationary phase (Pettit *et al*., 2014) we hypothesise a potential role for this sRNA in sporulation in *C. difficile* 630. Investigation into the expression of sRNAs in *C. difficile* 630 grown in faecal water enriched growth medium showed that sRNA expression changes in response to the presence of FW. The expression of two sRNAs complementary to the motility-associated genes *fliG* and *fliH*, was found to be decreased in FW indicating a possible positive regulatory role for the sRNA i.e. less sRNA expression equals less target gene expression. Previous research has found that sRNAs both positively and

negatively regulate the main genes responsible for flagellar synthesis in *E. coli*, *flhD* and *flhC* (De Lay and Gottesman, 2012; Bak *et al*., 2015) and motility, biofilm formation and virulence in *C. difficile* (Soutourina *et al.,* 2013). As we know that motility gene expression is reduced in *C. difficile* 630 grown in the presence of faecal water (Chapter 3; Ternan *et al*., 2018) it is reasonable to suggest that sRNAs may play a role in this. A sRNA and its complementary gene which encodes NorV (involved in the metabolism of nitric oxide, NO (Hutchings *et al*., 2002)) were both down in FW suggesting a positive regulatory relationship. NO is a key signalling molecule in the GI tract (Vermeiren *et al*., 2009) and research has shown that bilirubin acts as an endogenous scavenger of nitrous oxide (NO) within the gut thereby providing a protective mechanism to the presence of reactive species (Mancuso *et al*., 2006). We have therefore suggested that NorV, as a NO detoxifying enzyme, may be downregulated in the presence of bilirubin (within the faecal water milieu) as NO detoxification is no longer required by *C. difficile* in this growth environment (Kochar *et al*., 2011). The gene encoding XylR (a xylose repressor) was increased whereas its corresponding sRNA was decreased. This reduction in expression of the sRNA may have led to an increase in expression of the *xylR* gene as the sRNA's regulatory effect was absent. As XylR is involved in the regulation of xylose metabolism, it would suggest that this pathway is affected by the presence of faecal water possibly due to an adaptive response of the organism to other carbon sources and additional dietary metabolites within the FW. A possible regulatory role for sRNAs in sporulation within *C. difficile* 630 has also been identified where both sRNA and the gene encoding *SpoIIE* were found to be increased in FW. We know that expression of sporulation associated genes is increased in the FW environment from our previous work. Conversely, we saw a reduction in the number of spores present up to late-log

(although the changes were not statistically significant) (Chapter 3; Ternan *et al*., 2018) and we therefore suggest that sRNAs are involved in this process and may explain the discrepancy between the transcriptome and spores produced. This is the first investigation to show that expression levels of sRNA molecules change during growth in a biologically relevant growth medium and that these changes may in turn have significant effects on the expression of other genes potentially involved in virulence of the organism. With antibiotic resistance growing it has never been more important to uncover potential therapeutic and diagnostic targets to ensure the ability to treat infectious diseases in the future.

Although we have contributed significantly within this thesis to the current understanding of *C. difficile* 630 and many of the processes involved in pathogenesis, including motility and sporulation, we have also created many questions that we did not have the opportunity to pursue in this project. I believe further investigations could focus on the analysis of the transcriptome of the *spo0A* mutant strain grown in FW in order to further elucidate the mechanisms behind the large increase in sporulation found in our work in addition to further investigating the role of Spo0A beyond sporulation. Future studies to further elucidate the exact mechanisms behind sRNA regulation of genes in *C. difficile* could focus on the development of sRNA mutant strains in order to characterise the global effects of these important regulators. Confirmation of sRNA transcripts through Northern blotting would also add an extra layer of robustness in addition to the use of Western blotting to validate the protein product translated.

In conclusion, this thesis provides a deeper understanding of how the transcriptome of *C. difficile* 630 alters in response to faecal water and suggests that important molecular processes are not being picked up in studies that employ basic laboratory growth media. The impact of the faecal water growth environment on *C. difficile* is clear with significant expressional changes observed in sporulation, motility, virulence factors and an important role for regulatory sRNAs in these processes. We have provided a more realistic indication of how this organism may respond in the presence of human gut constituents whilst underscoring the regulatory effect FW has on gene expression in *C. difficile.*

# REFERENCES

Abt, M.C., McKenney, P.T., and Pamer, E.G. *Clostridium difficile* colitis: pathogenesis and host defence. *Nat Rev Microbiol*. (2016) 14(10): 609–20.

Advisory Committee on Antimicrobial Resistance and Healthcare Associated Infections: Updated guidance on the diagnosis and reporting of *Clostridium difficile*. United Kingdom: National Health Service, 2012. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/215135/dh_133016.pdf. Accessed Nov 2019.

Aguilar-Rodriguez, J., Sabater-Munoz, B., Montagud-Martinez, R., Berlanga, V., Alvarez-Ponce, D., Wagner, A. and Fares, M.A. The Molecular Chaperone DnaK Is a Source of Mutational Robustness. *Genome Biol. Evol*. (2016) 8(9): 2979-2991. doi:10.1093/gbe/evw176.

Akopian, D., Shen, K., Zhang, X. and Shan, S. Signal Recognition Particle: An essential protein targeting machine. *Annu Rev Biochem* (2013) 82: 693–721. doi:10.1146/annurev-biochem-072711-164732.

Al Assaad, R., Dakessian, A., Bachir, R., Bizri, A.R., and El Sayed, M. Significance of *Clostridium difficile* in community-acquired diarrhea in a tertiary care center in Lebanon. *Sci Rep*. (2020) 10(1):5678. doi:10.1038/s41598-020-62418-9.

Al-Hinai, M.A., Jones, S.W. and Papoutsakis, E.T. The *Clostridium* Sporulation Programs: Diversity and Preservation of Endospore Differentiation. *Microbiology and Molecular Biology Reviews* (2015) 79(1) 19-37. doi:10.1128/MMBR.00025-14.

Al-Hinai, M.A., Jones, S.W., and Papoutsakis, E.T. σK of *Clostridium acetobutylicum* is the first known sporulation-specific sigma factor with two developmentally separated roles, one early and one late in sporulation. *J Bacteriol*. (2014) 196(2): 287–99.

Allen, S.J., Wareham, K., Wang, D., Bradley, C., Sewell, B., Hutchings, H., Harris, W., Dhar, A., Brown, H., Foden, A., Gravenor, M.B., Mack, D., and Phillips, C.J. A high-dose preparation of l*actobacilli* and *bifidobacteria* in the prevention of antibiotic-associated and *Clostridium difficile* diarrhoea in older people admitted to hospital: a multicentre, randomised, double-blind, placebo-controlled, parallel arm trial (PLACIDE). *Health Technol. Assess*. (2013) 17: 1–140.

Alotaibi, B. The role of *Clostridium difficile* surface structures in virulence and the potential for host. PhD thesis. (2017) Ulster University.

Altuvia, S., Weinstein-Fischer, D., Zhang, A., Postow, L. and Storz, G. A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell* (1997) 90: 43–53. doi.org/10.1016/S0092-8674(00)80312-8

Anders, A. and Huber, W. Differential expression analysis for sequence count data. *BMC Genome Biology* (2010a)1:R106.

Andre, G., Even, S., Putzer, H., Burguiere, P., Croux, C., Danchin, A., Martin-Verstraete, I., and Soutourina, O. S-box and T-box riboswitches and antisense RNA control a sulfur metabolic operon of *Clostridium acetobutylicum*. *Nucleic Acids Res*. (2008) 36: 5955–5969.

Antunes, A., Martin-Verstraete, I., and Dupuy, B. CcpA-mediated repression of *Clostridium difficile* toxin gene expression. *Mol Microbiol*. (2011) 79(4):882–99.

Arora, G., Sajid, A., Virmani, R., Singhal, A., Kumar, C.M.S., Dhasmana, N., Khanna, T., Maji, A., Misra, R., Molle, V., Becher, D., Gerth, U.,. Mande, S.C. and Singh, Y. Ser/Thr protein kinase PrkC-mediated regulation of GroEL is critical for biofilm formation in *Bacillus anthracis*. *npj Biofilms and Microbiomes.* (2017) 3:7 doi:10.1038/s41522-017-0015-4.

Arriola, V., Tischendorf, J., Musuuza, J., Barker, A., Rozelle, J.W., and Safdar, N. Assessing the risk of hospital-acquired *Clostridium difficile* infection with proton pump inhibitor use: A meta-analysis. *Infect Control Hosp Epidemiol*. (2016) 37(12):1408-1417.

Asempa, T.E. and Nicolau, D.P. *Clostridium difficile* infection in the elderly: an update on management. *Clin Interv Aging*. (2017) 12:1799-1809. doi.org/10.2147/CIA.S149089.

Aubry, A., Hussack, G., Chen, W., Kuolee, R., Twine, S.M., Fulton, K.M., Foote, S., Carrillo, C.D., Tanha, J. and Logan, S.M. Modulation of toxin production by the flagellar regulon in *Clostridium difficile*. *Infect Immun* (2012) 80:3521–3532. doi:10.1128/IAI.00224-12.

Avila, M.B, Avila, N.P and Dupont, A.W. Recent Advances in the Diagnosis and Treatment of *Clostridium Difficile* Infection. *F1000Res*. (2016) 29;5: F1000 Faculty Rev-118. doi: 10.12688/f1000research.7109.1.

Awad, M.M., Johanesen, P.A., Carter, G.P., Rose, E., Lyras, D. *Clostridium difficile* virulence factors: Insights into an anaerobic spore-forming pathogen. *Gut Microbes* (2014) 5: 579–93.

Baban, S.T., Kuehne, S.A., Barketi-Klai, A., Cartman, S.T., Kelly, M.L., Hardie, K.R., Kansau, I., Collignon, A., Minton, N.P. The role of flagella in *Clostridium difficile* pathogenesis: comparison between a non-epidemic and an epidemic strain. *PLoS One* (2013) 8:e73026. doi:10.1371/journal.pone.0073026.

Bagasra, O., and Prilliman, K.R. RNA interference: the molecular immune system. *J. Mol. Histol.* (2004) 35 (6): 545-53.

Bagdasarian, N., Rao, K. and Malani, P.N. Diagnosis and treatment of *Clostridium difficile* in adults: a systematic review. *JAMA.* (2015) 313(4):398-408. doi:10.1001/jama.2014.17103.

Bak, G., Han, K., Kim, D. and Lee, Y. Roles of rpoS-activating small RNAs in pathways leading to acid resistance of *Escherichia coli*. *Microbiology Open* (2014) 3: 15–28. doi.org/10.1002/mbo3.143.

Bak, G., Lee, J., Suk, S., Kim, D., Lee, J.Y., Kim, K., Choi, B. and Lee, Y. Identification of novel sRNAs involved in biofilm formation, motility, and fimbriae formation in *Escherichia coli. Scientific Reports* (2015) 5: 15287. doi.org/10.1038/srep15287.

Baker, M.A. How biophysics may help us understand the flagellar motor of bacteria which cause infections. *Adv Exp Med Biol*. (2016) 915: 231-43.

Balsells, E., Shi, T., Leese, C., Lyell, I., Burrows, J., Wiuff, C., Campbell, H., Kyaw, M. H., and Nair, H. Global burden of *Clostridium difficile* infections: a systematic

review and meta-analysis. *Journal of global health* (2019) 9(1):010407. doi.org/10.7189/jogh.09.010407.

Barrick, J.E., and Breaker, R.R. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biology* (2007) 8: R239.

Bartlett J.G., Onderdonk, A.B., Cisneros, R.L., and Kasper, D.L. Clindamycin-associated colitis due to a toxin-producing species of *Clostridium* in hamsters. *J Infect Dis*. (1977) 136(5):701–705.

Bartlett, J.G. and Gerding, D.N. Clinical Recognition and Diagnosis of *Clostridium difficile* Infection. *J Clin Infect Dis*. (2008) 46: S12-S18.

Bartlett, J.G. *Clostridium difficile* infection: Historic review. *Anaerobe*. (2009) 15: 227-229.

Bartlett, J.G. Historical perspectives on studies of *Clostridium difficile* and *C. difficile* infection. *Clin Infect Dis*. (2008) 46: S4-S11. doi: 10.1086/521865.

Bartlett, J.G., Chang, T.W., Moon, N., and Onderdonk, A.B. Antibiotic-induced lethal enterocolitis in hamsters: studies with eleven agents and evidence to support the pathogenic role of toxin-producing *Clostridia*. *Am J Vet Res* (1978) 39: 1525-30.

Bartlett, J.G., Narrative Review: The New Epidemic of *Clostridium difficile* – Associated Enteric Disease. *Ann Intern Med*. (2006) 145: 758-764.

Batah, J., and Kansau, I. Intestinal epithelial cell response to *Clostridium difficile* flagella. *Methods Mol Biol*. (2016)1476: 103-16.

Bauer, M.P., Kuijper, E.J., and van Dissel, J.T. European Society of Clinical Microbiology and Infectious Diseases (ESCMID): treatment guidance document for *Clostridium difficile* infection (CDI). *Clin Microbiol Infect*. (2009) 15: 1067–1079.

Bendtsen, J.D., Kiemer, L., Fausbøll, A., and Brunak, S. Non-classical protein secretion in bacteria. *BMC Microbiology*. (2005) 5: 58.

Bernstein, E., Caudy, A., Hammond, S., and Hannon, G. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* (2001) 409 (6818): 363-366.

Best, E.L., Freeman, J., and Wilcox M.H. Models for the study of *Clostridium difficile* infection. *Gut Microbes*. (2012) 3(2):145-67.

Bidet, P., Lalande, V., Salauze, B., Burghoffer, B., Avesani, V., Delmée, M., Rossier, A., Barbut, F., and Petit, J.C. Comparison of PCR-ribotyping, arbitrarily primed PCR, and pulsed-field gel electrophoresis for typing *Clostridium difficile*. *J Clin Microbiol.* (2000) 38(7):2484-7. doi: 10.1128/JCM.38.7.2484-2487.2000.

Borriello, S.P., Ketley, J.M., Mitchell, T.J., Barclay, F.E., Welch, A.R., Price, A.B., *et al. Clostridium difficile*–a spectrum of virulence and analysis of putative virulence determinants in the hamster model of antibiotic-associated colitis. *J Med Microbiol* (1987) 24: 53-64. doi.org/10.1099/00222615-24-1-53.

Boudry, P., Gracia, C., Monot, M., Caillet, J., Saujet, L., Hajnsdorf, E., Dupuy, B., Martin-Verstraete, I., Soutourina, O. Pleiotropic Role of the RNA Chaperone Protein Hfq in the Human Pathogen *Clostridium difficile. Journal of Bacteriology* (2014) 196(18): 3234–3248. doi:10.1128/JB.01923-14.

Boyd, L.A., McCann, M.J., Hashim, Y., Bennett, R.N., Gill, C.I., and Rowland, I.R. Assessment of the anti-genotoxic, anti-proliferative, and anti-metastatic potential of crude watercress extract in human colon cancer cells. *Nutr Cancer*. (2006) 55(2): 232-41.

Bradshaw, M., Marshall, K.M., Heap, J.T., Tepp, W.H., Minton, N.P., and Johnson, E.A. Construction of a Nontoxigenic *Clostridium botulinum* Strain for Food Challenge Studies. *Appl. Environ. Microbiol*. (2010) 76: 387-93.

Brantl, S. Acting antisense: plasmid- and chromosome-encoded sRNAs from Gram-positive bacteria. *Future Microbiol.* (2012) 7(7):853-871. doi:10.2217/fmb.12.59.

Brazier, J.S. Typing of *Clostridium difficile*. *Clin Microbiol Infect*. (2001) 7(8): 428-431.

Brouwer, M., Roberts, A., Hussain, H., Williams, R., Allan, E. and Mullany, P. Horizontal gene transfer converts non-toxigenic *Clostridium difficile* strains into toxin producers. *Nat Commun* (2013) 4: 2601. doi.org/10.1038/ncomms3601.

Brown, A.W.W and Wilson, R.B. *Clostridium difficile* colitis and zoonotic origins-a narrative review. *Gastroenterol Rep (Oxf)*. (2018) 6(3):157-166. doi:10.1093/gastro/goy016

Brown, E.M., McDougall, G.J., Stewart, D., Pereira-Caro, G., González-Barrio, R., Allsopp, P., Magee, P., Crozier, A., Rowland, I., and Gill, C.I. Persistence of Anticancer Activity in Berry Extracts after Simulated Gastrointestinal Digestion and Colonic Fermentation. *PLOS ONE* (2012) doi.org/10.1371/journal.pone.0049740.

Browne, H.P., Forster, S.C, Anonye, B.O., Kumar, N., Neville, B.A., Stares, M.D., Goulding, D. and Lawley, T.D. Culturing of 'unculturable' human microbiota reveals novel taxa and extensive sporulation. *Nature* (2016) 533(7604): 543-546. doi:10.1038/nature17645.

Bruggemann, H., Baumer, S., Fricke, W.F., Wiezer, A., Liesegang, H., Decker, I., Herzberg, C., Martinez-Arias, R., Merkl, R., Henne, A., and Gottschalk, G. The genome sequence of *Clostridium tetani*, the causative agent of tetanus disease. *Proc Natl Acad Sci USA*. (2003) 100: 1316-1321.

Bruxelle, J.F., Tsapis, N., Hoys, S., Collignon, A., Janoir, C., Fattal, E., and Péchiné, S. Protection against *Clostridium difficile* infection in a hamster model by oral vaccination using flagellin FliC-loaded pectin beads. *Vaccine* (2018) 36(40): 6017-6021. doi.org/10.1016/j.vaccine.2018.08.013.

Burke, K.E., and Lamont, J.T. *Clostridium difficile* infection: a worldwide disease. *Gut Liver*. (2014) 8(1): 1-6.

Burnham, C.D. and Carroll, K.C. Diagnosis of *Clostridium difficile* Infection: an Ongoing Conundrum for Clinicians and for Clinical Laboratories. *Clin Microbiol Rev.* (2013) 26(3): 604–630. doi: 10.1128/CMR.00016-13.

Burns, D.A., Heap, J.T., and Minton, N.P. SleC is essential for germination of *Clostridium difficile* spores in nutrient-rich medium supplemented with the bile salt taurocholate. *J. Bacteriol*. (2010) 192: 657-64.

Bustin, S.A. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J. Mol. Endocrinol*. (2000) 25: 332-337.

Bustin, S.A. Why there is no link between measles virus and autism. In: Fitzgerald M (ed) Recent advances in autism spectrum disorders (2013) vol I, Intech. Rijeka, Croatia, pp 81–98.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J. and Wittwer, C.T. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem*. (2009) 55(4): 611-622.

Calloni, G., Chen, T., Schermann, S.M., Chang, H., Genevaux, P., Agostini, F., Tartaglia, G., Hayer-Hartl, M. and Hartl, F.U. DnaK Functions as a Central Hub in the *E. coli* Chaperone Network. *Cell Reports* (2012) 1(3) 251-264. doi.org/10.1016/j.celrep.2011.12.007.

Camiade, E., Peltier, J., Bourgeois, I., Couture-Tosi, E., Courtin, P., Antunes, A., Chapot-Chartier, M.P., Dupuy, B., and Pons, J.L. Characterization of Acp, a peptidoglycan hydrolase of *Clostridium perfringens* with N-acetylglucosaminidase activity that is implicated in cell separation and stress-induced autolysis. *J Bacteriol.* (2010) 192: 2373-84.

Cammarota, G., Ianiro, G., Kelly, C. R., Mullish, B. H., Allegretti, J. R., Kassam, Z., Putignani, L., Fischer, M., Keller, J. J., Costello, S. P., Sokol, H., *et al*. International consensus conference on stool banking for faecal microbiota transplantation in clinical practice. *Gut* (2019) 68(12): 2111–2121. doi.org/10.1136/gutjnl-2019-319548.

Cao, Y., Cao, Wu, J., Liu, Q., Zhao, Y., Ying, X., Cha, L., Wang, L., and Li, W. sRNATarBase: A comprehensive database of bacterial sRNA targets verified by experiments. *RNA* (2010) 16: 2051-2057.

Carter, G.P., Douce, G.R., Govind, R. Howarth, P.M., Mackin, K.E. *et al.* The anti-sigma factor TcdC modulates hypervirulence in an epidemic BI/NAP1/027 clinical isolate of *Clostridium difficile*. *PLoS Pathog* (2011) 7: e1002317.

Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, *et al.* The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res*. (2016) 44(D1):D471–80.

Chandrasekaran, R., and Lacy, D.B., The role of toxins in *Clostridium difficile* infection. *FEMS Microbiology Reviews* (2017) 41: 723–750. doi: 10.1093/femsre/fux048

Chastanet, A., Vitkup, D., Yuan, G.C., Norman, T.M., Liu, J.S., Losick, R.M. Broadly heterogeneous activation of the master regulator for sporulation in *Bacillus subtilis*. *Proc Natl Acad Sci USA* (2010) 107(18): 8486-8491.

Chen, Y., Indurthi, D. C., Jones, S. W., Papoutsakis, E. T. Small RNAs in the genus *Clostridium. MBio*. (2011) 2(1): doi: 10.1128/mBio.00340-10

Christiansen, J.K., Larsen, M.H., Ingmer, H., Sogaard- Andersen, L., and Kallipolitis, B.H. The RNA-binding protein Hfq of *Listeria monocytogenes*: role in stress tolerance and virulence. *J Bacteriol.* (2004) 186: 3355-3362.

Claesson, M.J., Jeffery, I.B., Conde, S., Power, S.E., O'Connor, E.M., Cusack, S., *et al*. Gut microbiota composition correlates with diet and health in the elderly. *Nature*. (2012) 488(7410): 178–84.

Clements, A.C., Magalhaes, R.J., Tatem, A.J., Paterson, D.L., Riley, T.V. *Clostridium difficile* PCR ribotype 027: assessing the risks of further worldwide spread. *Lancet Infect Dis* (2010) 10: 395-404.

Collery, M.M., Kuehne, S.A., McBride, S.M., Kelly, M.L., Monot, M., Cockayne, A., Dupuy, B., and Minton, N.P. What's a SNP between friends: The influence of single nucleotide polymorphisms on virulence and phenotypes of *Clostridium difficile* strain 630 and derivatives. *Virulence.* (2017) 18:8(6):767-781. doi:10.1080/21505594.2016.1237333.

Commission for Healthcare Audit and Inspection. Investigation into outbreaks of *Clostridium difficile* at Stoke Mandeville Hospital, Buckinghamshire Hospitals NHS Trust. Commission for Healthcare Audit and Inspection, (2006) London, United Kingdom. www.cqc.org.uk/_db/_documents/Stoke_Mandeville.pdf.

Cooksley, C.M., Davis, I.J., Winzer, K., Chan, W.C., Peck, M.W., and Minton, N.P. Regulation of neurotoxin production and sporulation by a putative agrBD signaling system in proteolytic *Clostridium botulinum. Appl. Environ. Microbiol*. (2010) 76: 4448-60.

Crobach M.J, Planche, T., Eckert, C., Barbut, F., Terveer, E.M., Dekkers, O.M., *et al*. European Society of Clinical Microbiology and Infectious Diseases: update of the diagnostic guidance document for *Clostridium difficile* infection. *Clin Microbiol Infect*. (2016) 22:(4): S63–81.

Crobach, M.J., Dekkers, O.M., Wilcox, M.H., and Kuijper, E.J. European Society of Clinical Microbiology and Infectious Diseases (ESCMID): Data review and recommendations for diagnosing *Clostridium difficile*-infection (CDI). *Clin Microbiol Infect.* (2009) 15: 1053-1066.

Curry, S.R. *Clostridium difficile*. *Clin Lab Med*. (2010) 30(1): 329–342. doi:10.1016/j.cll.2010.04.001.

Czepiel, J., Dróżdż, M., Pituch, H., Kuijper, E.J., Perucki, W., Mielimonka, A., Goldman, S., Wultańska, D. Garlicki, A. and Biesiada, G. *Clostridium difficile* infection: review. *European Journal of Clinical Microbiology & Infectious Diseases* (2019) 38:1211–1221. doi.org/10.1007/s10096-019-03539-6.

Daniela, E., Sara, S., Marcella, M., Giovanni, A., Meynier, A., Sophie, V., *et al*. Fecal water genotoxicity in healthy free–living young Italian people. *Food Chem Toxicol*. (2014) 64: 104-9.

Dapa, T., Leuzzi, R., Ng, Y.K., Baban, S.T., Adamo, R., Kuehne, S.A., Scarselli, M., Minton, N.P., Serruto, D., Unnikrishnan, M. Multiple factors modulate biofilm formation by the anaerobic pathogen *Clostridium difficile*. *J Bacteriol* (2013) 195(3): 545–555.

Davies, A.H., Roberts, A.K., Shone, C.C. and Acharya, K.R. Super toxins from a super bug: structure and function of *Clostridium difficile* toxins. *Biochem J.* (2011) 436(3):517-26. doi:10.1042/BJ20110106. PMID: 21615333.

Dawson, L.F., Valiente, E., and Wren, B.W. *Clostridium difficile* – A continually evolving and problematic pathogen. *Infection, Genetics and Evolution*. (2009) 9: 1410-1417.

Dawson, L.F., Valiente, E., Faulds-Pain, A., Donahue, E.H., Wren, B.W. Characterisation of *Clostridium difficile* biofilm formation, a role for Spo0A. *PloS One* (2012) 7(12): e50527.

De Lay, N. and Gottesman, S. A complex network of small non-coding RNAs regulate motility in *Escherichia coli. Molecular Microbiology* (2012) 86(3): 524–538.

Deakin, L., J., C., S., Fagan, R., P., Dawson, L., F., Pickard, D., J., West, M., R., Wren, B., W., Fairweather, N., F., Dougan, G., and Lawley, T., D. The *Clostridium difficile spo0A* gene is a persistence and transmission factor. *Infection and Immunity.* (2012) 80: 2704-2711.

Debast, S. B., Bauer, M. P. & Kuijper, E. J. European Society of Clinical Microbiology and Infectious Diseases: update of the treatment guidance document for *Clostridium difficile* infection. *Clin. Microbiol. Infect*. (2014) 20(2): 1–26.

Deligianni, E., Pattison, S., Berrar, D., Ternan, N.G., Haylock, R.W., Moore, J.E., *et al*. *Pseudomonas aeruginosa* cystic fibrosis isolates of similar RAPD genotype exhibit diversity in biofilm forming ability *in vitro*. *BMC Microbiol*. (2010) 10: 38.

Delihas, N. and Forst, S. Micf: an antisense RNA gene involved in response of *Escherichia coli* to global stress factors. *J. Mol. Biol*. (2001) 313: 1–12 doi.org/10.1006/jmbi.2001.5029.

Dembek, M. Whole-genome analysis of sporulation and germination in *Clostridium difficile*. https://spiral.imperial.ac.uk/handle/10044/1/38630 (2014). Accessed 2 Nov 2017.

Denève, C., Deloménie, C., Barc, M., Collignon, A. and Janoir, C. Antibiotics involved in *Clostridium difficile*-associated disease increase colonization factor gene expression. *J Med Microbiol* (2008) 57:723-728.

Denève, C., Janoir, C., Poilane, I., Fantinato, C., and Collignon, A. New trends in *Clostridium difficile* virulence and pathogenesis. *Int J Antimicro Ag.* (2009) 33(S1): S24- S28.

Deng, H., Yang, S., Zhang, Y., Qian, K., Zhang, Z., Liu, Y., Wang, Y., Bai, Y., Fan, H., Zhao, X. and Zhi, F. *Bacteroides fragilis* Prevents *Clostridium difficile* Infection in a Mouse Model by Restoring Gut Barrier and Microbiome Regulation. *Front. Microbiol.* (2018) 9: 2976. doi: 10.3389/fmicb.2018.02976

Deutscher, J., Aké, F.M., Derkaoui, M., Zébré, A.C., Cao, T.N., Bouraoui, H., *et al*. The bacterial phosphoenolpyruvate:carbohydrate phosphotransferase system: regulation by protein phosphorylation and phosphorylation-dependent protein–protein interactions. *Microbiol Mol Biol Rev*. (2014) 78(2):231–56.

Di Bella, S., Ascenzi, P., Siarakas, S., Petrosillo, N. and Di Masi, A. *Clostridium difficile* Toxins A and B: Insights into Pathogenic Properties and Extraintestinal Effects. *Toxins* (2016) 8: 134; doi:10.3390/toxins8050134.

Dingle, K.E, Didelot, X., Ansari, M.A., Eyre, D.W., Vaughan, A., Griffiths, D., Ip, C.L., Batty, E.M., Golubchik, T., Bowden, R., Jolley, K.A., Hood, D.W., Fawley, W.N., Walker, A.S., Peto, T.E., Wilcox, M.H., and Crook, D.W. Recombinational

switching of the *Clostridium difficile* S-layer and a novel glycosylation gene cluster revealed by large scale whole genome sequencing. *J. Infect. Dis*. (2013) 207:675– 686.

Dingle, T.C., Mulvey, G.L. and Armstrong, G.D. Mutagenic analysis of the *Clostridium difficile* flagellar proteins, FliC and FliD, and their contribution to virulence in hamsters. *Infect Immun* (2011) 79:4061–4067. doi:10.1128/IAI.05305-11.

Dodd, M.S., Papineau, D., Grenne, T., Slack, J.F., Rittner, M., Pirajno, F., O'Neil, J., Little, C.T.S. Evidence for early life in Earth's oldest hydrothermal vent precipitates. *Nature*. (2017) 543 (7643): 60–64. doi:10.1038/nature21377.

Dong, H., Zhang, Y., Dai, Z., and Li, Y. Engineering clostridium strain to accept unmethylated DNA. *PLoS One*. (2010) 5: e9038.

Draghici, S., Khatri, P., Eklund, A.C., and Szallasi, Z. Reliability and reproducibility issues in DNA microarray measurements. *Trends Genet*. (2006) 22(2): 101-109.

Dupuy, B., Govind, R., Antunes, A., and Matamouros, S. *Clostridium difficile* toxin synthesis is negatively regulated by TcdC. *Journal of Medical Microbiology* (2008) 57: 685–689.

Edgar, R., Domrachev, M., and Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* (2002) 30(1): 207–10.

Edwards, A.N., Nawrocki, K.L., and McBride, S.M. Conserved oligopeptide permeases modulate sporulation initiation in *Clostridium difficile*. *Infect Immun*. (2014) 82(10):4276–91.

Edwards, A.N., Tamayo, R., and McBride, S.M. A Novel Regulator Controls *Clostridium difficile* Sporulation, Motility and Toxin Production. *Mol Microbiol*. (2016) 100(6): 954-971. doi:10.1111/mmi.13361.

Eid, N., Osmanova, H., Natchez, C., Walton, G., Costabile, A., Gibson, G., *et al*. Impact of palm date consumption on microbiota growth and large intestinal health: a randomised, controlled, cross-over, human intervention study. *Br J Nutr*. (2015) 28;114(8):1226-36.

Elliott, B., Androga, G., & Daniel, K., and Thomas, R. *Clostridium difficile* infection: Evolution, phylogeny and molecular epidemiology. *Infection Genetics and Evolution* (2016) 49. 1-11. 10.1016/j.meegid.2016.12.018.

Elsayed, S., and Zhang, K. Bacteremia caused by *Clostridium intestinale*. *J Clin Microbiol*. (2004) 43(4): 2018-2020.

Emerson, J. E., Stabler, R. A., Wren, B. W. and N. F. Fairweather. Microarray analysis of the transcriptional responses of *Clostridium difficile* to environmental and antibiotic stress. *J Med Microbiol*. (2008) 57:757-64.

Emerson, J.E., Reynolds, C.B., Fagan, R.P., Shaw, H.A., Goulding, D., and Fairweather, N.F. A novel genetic switch controls phase variable expression of CwpV, a *Clostridium difficile* cell wall protein. *Mol. Microbiol*. (2009) 74: 541-556.

Evans, L.D., Hughes, C., and Fraser, G.M. Building a flagellum outside the bacterial cell. *Trends Microbiol*. (2014) 22(10): 566-72.

Ezeji, T.C., Qureshi, N. Blaschek, H.P. Production of acetone, butanol and ethanol by *Clostridium beijerinckii* BA101 and in situ recovery by gas stripping. *World Journal of Microbiology and Biotechnology* (2003) 19(6): 595–603.

Fagan, R. P., Albesa-Jové, D., Qazi, O., Svergun, D. I., Brown, K. a. and Fairweather, N. F. (2009) 'Structural insights into the molecular organization of the S-layer from Clostridium difficile', *Molecular Microbiology*, 71(5), pp. 1308–1322.

Faron, M.L., Ledeboer, N.A., Buchan, B.W. Resistance mechanisms, epidemiology, and approaches to screening for vancomycin-resistant *Enterococcus* in the health care setting. *J Clin Microbiol* (2016) 54: 2436 –2447. doi:10.1128/JCM.00211-16.

Fenner, L., Widmer, A.F., Goy, G., Rudin, S., and Reno Frei, R. Rapid and Reliable Diagnostic Algorithm for Detection of *Clostridium difficile*. *J Clin Microbiol*. (2008) 46(1): 328-330.

Ferreyra, J.A., Wu, K.J., Hryckowian, A.J., Bouley, D.M., Weimer, B.C., Sonnenburg, J.L. Gut microbiota-produced succinate promotes *C. difficile* infection after antibiotic treatment or motility disturbance. *Cell Host Microbe*. (2014) 16: 770-7.

Fimlaid, K. A., and Shen, A. Diverse mechanisms regulate sporulation sigma factor activity in the Firmicutes. *Curr. Opin. Microbiol*. (2015) 24: 88–95. doi: 10.1016/j.mib.2015.01.006

Fimlaid, K.A., Bond, J.P., Schutz, K.C., Putnam, E.E., Leung, J.M., Lawley, T.D, *et al*. Global Analysis of the Sporulation Pathway of *Clostridium difficile*. *PLoS Genet* (2013) 9(8): e1003660. doi.org/10.1371/journal.pgen.1003660.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* (1998) 391: 806-811.

Franze de Fernandez, M.T., Eoyang, L. and August, J.T. Factor fraction required for the synthesis of bacteriophage Qbeta-RNA. *Nature*. (1968) 219(5154):588-90. doi: 10.1038/219588a0.

Freedberg, D.E., and Abrams, J.A. *Clostridium difficile* infection in the community: are proton pump inhibitors to blame? *World J Gastroenterol.*(2013) 28;19(40): 6710-3.

Freeman, J., Bauer, M.P., Baines, S. D. Corver, J. Fawley, W. N., Goorhuis, B., Kuijper, E. J. and Wilcox, M. H. The Changing Epidemiology of *Clostridium difficile* Infections. *Clinical Microbiology Reviews* (2010) 23(3) 529–549. doi:10.1128/CMR.00082-09.

Furuya-Kanamori, L., Marquess, J., Yakob, L., Riley, T. V., Paterson, D. L., Foster, N. F., Huber, C. A., and Clements, A. C. Asymptomatic Clostridium difficile colonization: epidemiology and clinical implications. *BMC infectious diseases* (2015) 15: 516. doi.org/10.1186/s12879-015-1258-4.

Galdys, A.L., Nelson, J.S., Shutt, K.A., *et al.* Prevalence and duration of asymptomatic *Clostridium difficile* carriage among healthy subjects in Pittsburgh, Pennsylvania. *J Clin Microbiol*. (2014) 52(7): 2406-2409. doi:10.1128/JCM.00222-14.

Galperin, M.Y. Genome Diversity of Spore-Forming *Firmicutes*. *Microbiol. Spectr.* (2013) 1(2): TBS-0015-2012. doi:10.1128/ microbiolspectrum.TBS-0015-2012.

Gao, X. W., Mubasher, M., Fang, C. Y., Reifer, C. and Miller, L. E. Dose-response efficacy of a proprietary probiotic formula of *Lactobacillus acidophilus* CL1285 and *Lactobacillus casei* LBC80R for antibiotic-associated diarrhea and *Clostridium difficile*-associated diarrhea prophylaxis in adult patients. *Am. J. Gastroenterol*. (2010) 105: 1636–1641.

Garzon, R., Calin, G.A., and Croce, C.M. MicroRNAs in cancer. *Annu Rev Med.* (2009) 60: 167–179.

Genth, H., Dreger, S.C., Huelsenbeck, J., and Just, I. *Clostridium difficile* toxins: More than mere inhibitors of Rho proteins. *Int J Biochem Cell Biol*. (2008) 40: 592– 597.

Georgescauld, F., Popova, K., Gupta, A.J., Bracher, A., Engen, J.R., Hayer-Hartl, M., and Hartl, F.U. GroEL/ES Chaperonin Modulates the Mechanism and Accelerates the

Rate of TIM-Barrel Domain Folding. *Cell*. (2014) 157(4): 922–934. doi:10.1016/j.cell.2014.03.038.

Gerding, D. N., File, T. M. and McDonald, L. C. Diagnosis and treatment of *Clostridium difficile* infection. *Infect. Dis. Clin. Pract.* (2016) 24: 3–10.

Gerding, D.N., Muto, C.A., and Owens, R.C., (2008) Measures to Control and Prevent Clostridium difficile Infection. Clin Infect Dis. 46 (1): S43-S49.

Gill, C.I., Heavey, P., McConville, E., *et al.* Effect of faecal water on an *in vitro* model of colonic mucosal barrier function. *Nutr Cancer*. (2007) 57(1):59-65. doi:10.1080/01635580701268220.

Gill, C.I., McDougall, G.J., Glidewell, S., Stewart, D., Shen, Q., Tuohy, K., *et al*. Profiling of phenols in human fecal water after raspberry supplementation. *J Agric Food Chem*. (2010) 58(19): 10389-95.

Gilligan, P.H. Optimizing the Laboratory Diagnosis of *Clostridium difficile* Infection. *Clin Lab Med*. (2015) 35(2):299–312. doi:10.1016/j.cll.2015.02.003.

Gottesman, S. Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends. Genet.* (2005) 21: 399–404.

Govind, R., and Dupuy, B. Secretion of *Clostridium difficile* toxins A and B requires the holin-like protein TcdE. *PLoS Pathog* (2012) 8:e1002727.

Green, N.J., Grundy, F.J., and Henkin, T.M. The T box mechanism: tRNA as a regulatory molecule. *FEBS Lett*. (2010) 584(2): 318–324. doi:10.1016/j.febslet.2009.11.056.

Griffiths, D., Fawley, W., Kachrimanidou, M., Bowden, R., Crook, D.W., Fung, R., Golubchik, T., Harding, R.M., Jeffery, K.J., Jolley, K.A., Kirton, R., Peto, T.E., Rees, G., Stoesser, N., Vaughan, A., Walker, A.S., Young, B.C., Wilcox, M., and Dingle, K.E. Multilocus sequence typing of *Clostridium difficile*. *J Clin Microbiol*. (2010) 48: 770- 778.

Grunberg-Manago, M. Regulation of the expression of aminoacyl-tRNA synthetases and translation factors In: Neidhardt F.C., Curtiss, R., Ingraham, J.L., Lin, E.C., Low, K.B., Magasanik. B., Reznikoff, W.S., Riley, M., Schaechter, M., Umbarger, H.E., eds. *Escherichia coli* and *Salmonella*: cellular and molecular biology Washington, D.C.: *American Society for Microbiology* (1996) 1432–1457.

Grundy, F.J., Rollins, S.M., and Henkin, T.M. Interaction between the acceptor end of tRNA and the T-box stimulates antitermination in the Bacillus subtilis tyrS gene: a new role for the discriminator base. *J Bacteriol*. (1994) 176(15): 4518–4526.

Gurtler, V., Wilson, V.A. and Mayall, B.C. Classification of medically important clostridia using restriction endonuclease site differences of PCR-amplified 16s rDNA. *Journal of General Microbiology* (1991) 137: 2673-2679.

Hannon, G. RNA interference. *Nature* (2002) 418: 244-251.

Hardy, K., Manzoor, S., Marriott, C., Parsons, H., Waddington, C. *et al.* Utilizing rapid multiple-locus variable-number tandem-repeat analysis typing to aid control of hospital-acquired *Clostridium difficile* infection: a multicenter study. *J. Clin. Microbiol*. (2012) 50:3244 –3248.

He, M., Sebaihia, M., Lawley, T.D., Stabler, R.A., Dawson, L.F. *et al.* Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc. Natl. Acad. Sci. USA*. (2010) 107: 7527-7532.

Heap, J., T., Cartman, S., T., Kuehne, S., A., Cooksley, C., and Minton, N., P. ClosTron-targeted mutagenesis. *Methods in Molecular Biology*. (2010) 646: 165-82.

Heap, J., T., Pennington, O., J., Cartman, S., T., Carter, G., P., and Minton, N., P. The ClosTron: A universal gene knock-out system for the genus *Clostridium*. *Journal of Microbiological Methods*. (2007) 70: 452-464.

Heinlen, L., and Ballard, J.D. *Clostridium difficile* Infection. Am J Med Sci. (2010) 340(3): 247–252. doi:10.1097/MAJ.0b013e3181e939d8.

Hemme, C.L., Mouttaki, H., Lee, Y., Zhang, G., Goodwin, L. *et al*. Sequencing of Multiple Clostridial Genomes Related to Biomass Conversion and Biofuel Production. *American Society for Microbiology*. (2010) 192(24): 6494-6496.

Henkin, T.M., Glass, B.L., and Grundy, F.J. Analysis of the *Bacillus subtilis tyrS* gene: Conservation of a regulatory sequence in multiple tRNA synthetase genes. *J. Bacteriol*. (1992) 174(4):1299–1306. [PubMed: 1735721]

Hennequin, C., Collignon, A., and Karjalainen, T. Analysis of expression of GroEL (Hsp60) of *Clostridium difficile* in response to stress. *Microbial Pathogenesis*. (2001a) 31: 255-260.

Hennequin, C., Janoir, C., Barc, M.C., Collingnon, A., and Karjalainen, T. Identification and characterization of a fibronectin-binding protein from *Clostridium difficile. Microbiology*. (2003) 149: 2779-2787.

Hennequin, C., Porcheray, F., Waligora-Dupriet, A., Collingnon, A., Barc, M., Bourlioux, P., and Karjalainen, T. GroEL (Hsp60) of *Clostridium difficile* is involved in cell adherence. *Microbiology*. (2001b) 147: 87-96.

Heo, G.Y., Kim, W.C., Joo, G.J., Kwak, Y.Y., Shin, J.H., Roh, D.H., Park, H.D., Rhee, I.K. Deletion of xylR Gene Enhances Expression of Xylose Isomerase in *Streptomyces lividans* TK24. *J. Microbiol. Biotechnol*. (2008) 18(5): 837-844.

Hsu, J., Abad, C., Dinh, M., and Safdar, N. Prevention of Endemic Healthcare-Associated *Clostridium difficile* Infection: Reviewing the Evidence. *Am J Gastroenterol.* (2010) 105: 2327-2339.

Huber, C., A., Foster, N., F., Riley, T., V., and Paterson, D., L. Challenges for Standardization of *Clostridium difficile* Typing Methods. *Journal of Clinical Microbiology* (2013) 51:2810-2814.

Humphries, R.M. and Linscott, A.J. Laboratory diagnosis of bacterial gastroenteritis. *Clin Microbiol Rev* (2015) 28:3–31. doi:10.1128/CMR.00073-14.

Hussain, H., A., Roberts, A., P., and Mullany, P. Generation of an erythromycin-sensitive derivative of *Clostridium difficile* strain 630 (630Deltaerm) and

demonstration that the conjugative transposon Tn916DeltaE enters the genome of this strain at multiple sites. *Journal of Medical Microbiology* (2005) 54: 137-141.

Hutchings, M.I., Mandhana, N. and Spiro, S. The NorR Protein of Escherichia coli Activates Expression of the Flavorubredoxin Gene norV in Response to Reactive Nitrogen Species. *Journal of Bacteriology* (2002) 184(16): 4640–4643. doi:10.1128/JB.184.16.4640–4643.2002.

Ignatius, R., Neuber, R., Kietzmann, H., Berg, C., Wenzel, T., Fuhrmann, J., and Müller, M. *Clostridioides difficile* in Outpatients: Application of a Diagnostic Algorithm Recommended by the European Society of Clinical Microbiology and Infectious Diseases. *European Journal of Microbiology and Immunology*. (2019)9:3: 88–90. DOI: 10.1556/1886.2019.00017.

Itikawa, H. and Ryu, J. Isolation and characterization of a temperature-sensitive *dnaK* mutant of *Escherichia coli* B. *J. Bacteriol*. (1979) 138: 339–344.

Jackson, M., Olefson, S., Machan, J.T., and Kelly, C.R. A High Rate of Alternative Diagnoses in Patients Referred for Presumed *Clostridium difficile* Infection. *J Clin Gastroenterol*. (2016) 50(9):742-746. doi:10.1097/MCG.0000000000000447.

Jain, S. A functional genomics investigation of clinically relevant heat-stress in Clostridium difficile strain 630. PhD Thesis. (2010) University of Ulster.

Jain, S., Graham, C., Graham, R.L., McMullan, G., and Ternan, N.G. Quantitative proteomic analysis of the heat stress response in *Clostridium difficile* strain 630. *J Proteome Res*. (2011) 10(9): 3880-90.

Jain, S., Smyth, D., O'Hagan, B., Heap, J., McMullan, G., Minton, N., and Ternan, N. Inactivation of the dnaK gene in *Clostridium difficile* 630 Delta erm yields a temperature-sensitive phenotype and increases biofilm-forming ability. *Scientific Reports* (2017) 7:17522. doi.org/10.1038/s41598-017-17583-9.

Janarthanan, S., Ditah, I., Adler, D.G., and Ehrinpreis, M.N. *Clostridium difficile*-associated diarrhea and proton pump inhibitor therapy: a meta-analysis. *Am J Gastroenterol*. (2012) 107:1001–1010

Janoir, C., Pechine, S., Grosdidier, C., and Collignon, A. Cwp84, a surface- associated protein of *Clostridium difficile*, is a cysteine protease with degrading activity on extracellular matrix proteins. *J Bacteriol*. (2007) 189: 7174-7180.

Jarrad, A.M., Karoli, T., Blaskovich, M.A., Lyras, D., and Cooper, M.A. *Clostridium difficile* drug pipeline: challenges in discovery and development of new agents. *J. Med. Chem*. (2015) 58(13): 5164-85.

Johnston, B. C., Ma, S.S., Goldenberg, J.Z., Thorlund, K., Vandvik, P.O., Loeb, M., and Guyatt, G.H. Probiotics for the prevention of *Clostridium difficile*-associated diarrhea: a systematic review and meta-analysis. *Ann. Intern. Med*. (2012) 157: 878–888.

Joseph, R.C., Kim, N.M. and Sandoval, N.R. Recent Developments of the Synthetic Biology Toolkit for *Clostridium. Front. Microbiol.* (2018) 9:154. doi: 10.3389/fmicb.2018.00154

Justice, S.S., Hung, C., Theriot, J.A., Fletcher, D.A., Anderson, G.G., Footer, M.J., *et al*. Differentiation and developmental pathways of uropathogenic *Escherichia coli* in urinary tract pathogenesis. *Proc Natl Acad Sci* (2004) 101: 1333e8.

Justice, S.S., Hunstad, D., Cegelski, L. *et al*. Morphological plasticity as a bacterial survival strategy. *Nat Rev Microbiol* (2008) 6: 162–168 doi.org/10.1038/nrmicro1820.

Kalia, V.C., Mukherjee, T., Bhushan, A., Joshi, J., Shankar, P., and Huma, N. Analysis of the unexplored features of rrs (16S rDNA) of the Genus *Clostridium*. *BMC Genomics* (2011) 12:18.

Karp, P.D., Latendresse, M., Paley, S.M., Krummenacker, M., Ong, Q.D., Billington, R., *et al*. Pathway Tools version 19.0: Integrated Software for Pathway/Genome Informatics and Systems Biology. *Brief Bioinform*. (2016) 17(5): 877-90.

Kassam, Z., Lee, C. H., Yuan, Y. and Hunt, R. H. Fecal microbiota transplantation for *Clostridium difficile* infection: systematic review and meta-analysis. *Am. J. Gastroenterol.* (2013) 108: 500–508.

Kelly, C.P. and LaMont, J.T. *Clostridium difficile* Infection. *Annu Rev Med*. (1998). 49: 375-390.

Khan, F.Y. and Elzouki, A.N. *Clostridium difficile* infection: a review of the literature. *Asian Pac J Trop Med* (2014) 7(Suppl 1): S6-S13.

Khanna, S., Pardi, D.S., Aronson, S.L., Kammer, P.P., Orenstein, R., St Sauver, J.L., Harmsen, W.S., Zinsmeister, A.R. The epidemiology of community-acquired *Clostridium difficile* infection: a population-based study. *Am J Gastroenterol*. (2012) 107(1):89–95.

Kim, J. N. and Breaker, R. R. Purine sensing by riboswitches. *Biology of the Cell* (2008) 100 (1), 1-11.

Kirby, J.M., Ahern, H., Roberts, A.K., Kumar, V., Freeman, Z., Acharya, K.R., and Shone, C.C. Cwp84, a surface-associated cysteine protease, plays a role in the maturation of the surface layer of *Clostridium difficile. J. Biol. Chem*. (2009) 284: 34666-34673.

Kochan, T.J., Foley, M.H., Shoshiev, M.S., Somers, M.J., Carlson, P.E., and Hanna, P.C. Updates to *Clostridium difficile* Spore Germination. *J Bacteriol* (2018) 200(16): e00218-18. doi.org/10.1128/JB.00218-18.

Kochar, N.I., Chandewal, A.V., Bakal, R.L., and Kochar, P.N. Nitric Oxide and the Gastrointestinal Tract. *International Journal of Pharmacology* (2011) 7 (1): 31-39. doi:10.3923/ijp.2011.31.39.

Kovacs-Simon, A., Leuzzi, R., Kasendra, M., Minton, N., Titball, R.W., and Michell, S.L. Lipoprotein CD0873 is a novel adhesin of *Clostridium difficile*. *J Infect Dis*. (2014) 210(2): 274–84.

Kroos, L., Kuspa, A., and Kaiser, D. A global analysis of developmentally regulated genes in *Myxococcus xanthus*. *Developmental Biology*. (1986) 117 (1): 252-266. doi:10.1016/0012-1606(86)90368-4.

Kuehne, S.A. and Minton, N.P. ClosTron-mediated engineering of Clostridium. *Bioengineered*. (2012) 3(4):247-54. doi:10.4161/bioe.21004.

Kuehne, S.A., Heap, J.T., Cooksley, C.M., Cartman, S.T., and Minton, N.P. ClosTron-Mediated Engineering of *Clostridium*. In: Williams J. (eds) Strain Engineering. Methods in Molecular Biology (Methods and Protocols) (2011) vol 765. Humana Press.

Kuijper, E.J., van den Berg, R.J., and Brazier, J.S. Comparison of molecular typing methods applied to *Clostridium difficile. Methods Mol. Biol*. (2009) 551:159 –171.

Kukurba, K.R., and Montgomery, S.B. RNA Sequencing and Analysis. *Cold Spring Harbor Protocols*. (2015) 11: 951-69. doi:10.1101/pdb.top084970

Kwok, C.S., Arthur, A.K., Anibueze, C.I., Singh, S., Cavallazzi, R., and Loke, Y.K. Risk of *Clostridium difficile* infection with acid suppressing drugs and antibiotics: meta-analysis. *Am J Gastroenterol*. (2012) 107:1011–1019.

Kyne, L., C. Merry, B., O'Connell, C., Keane, and D., O'Neill. Community-acquired *Clostridium difficile* infection. *J. Infect*. (1998) 6:287–288.

Landan, G., and Graur, D. Heads or tails: A simple reliability check for multiple sequence alignments. *Mol Biol Evol*. (2007) 24: 1380–1383. doi.org/10.1093/molbev/msm060.

Lawley, T.D., Croucher, N.J., Yu, L., Clare, S., Sebaihia, M., Goulding, D., *et al*. Proteomic and genomic characterization of highly infectious Clostridium difficile 630 spores. *J Bacteriol*. (2009) 191(17):5377–86

Lawson, P.A., Citron, D.M., Tyrrell, K.L. and Finegold, S.M. Reclassification of *Clostridium difficile* as *Clostridioides difficile* (Hall and O'Toole 1935) Prévot 1938. *Anaerobe* (2016) 40:95-99. doi:10.1016/j.anaerobe.2016.06.008.

Lease, R.A., Smith, D., McDonough, K. and Belfort, M. The small noncoding DsrA RNA is an acid resistance regulator in *Escherichia coli. J. Bacteriol*. (2004) 186: 6179–6185. doi.org/10.1128/JB.186.18.6179-6185.2004.

Lechat P, Hummel L, Rousseau S, and Moszer I. GenoList: an integrated environment for comparative analysis of microbial genomes. *Nucleic Acids Res*. (2008) 36(Database issue):D469–74.

Leffler, D. A. and Lamont, J. T. *Clostridium difficile* infection. *N. Engl. J. Med.* (2015) 372: 1539–1548.

Lemée, L., Dhalluin, A., Pestel-Caron, M., Lemeland, J. and Pons, J. Multilocus sequence typing analysis of human and animal *Clostridium difficile* isolates of various toxigenic types. *J Clin Microbiol*. (2004) 42:2609-2617.

Li, C., Parker, A., Menocal, E., Xiang, S., Borodyansky, L., and Fruehauf, J. Delivery of RNA interference. *Cell Cycle* (2006) 5 (18): 2103-2109.

Li, J., Wang, Y., Zhang, C.Y., Zhang, W.Y., Jiang, D.M., Wu, Z.H., Liu, H., and Li, Y.Z. *Myxococcus xanthus* viability depends on groEL supplied by either of two genes, but the paralogs have different functions during heat shock, predation, and development. *J Bacteriol*. (2010) 192: 1875-188.

Li, L., Huang, D., Cheung, M.K., Nong, W., Huang, Q., and Kwan, H.S. BSRD: a repository for bacterial small regulatory RNA. *Nucleic Acids Research*. (2013) 41(D1): D233-D238. doi.org/10.1093/nar/gks1264.

Liberek, K., Lewandowska, A., and Zietkiewicz, S. Chaperones in control of protein disaggregation. *EMBO J.* (2008) 27: 328-335.

Lindström, M., and Korkeala, H. Laboratory diagnostics of botulism. *Clin. Microbiol. Rev*. (2006) 19:298-314.

Macfarlane, G.T., Macfarlane, S., and Gibson, G.R. Validation of a Three-Stage Compound Continuous Culture System for Investigating the Effect of Retention Time on the Ecology and Metabolism of Bacteria in the Human Colon. *Microb Ecol* (1998) 35:180-7.

Mack, G.S. "MicroRNA gets down to business". *Nature Biotechnology* (2007). 25 (6): 631–8. doi:10.1038/nbt0607-631.

Mackin, K.E., Carter, G.P., Howarth, P., Rood, J.I., Lyras, D. Spo0A differentially regulates toxin production in evolutionarily diverse strains of *Clostridium difficile*. *PloS One* (2013) 8(11): e79666.

Majdalani, N., Chen, S.A., Murrow, J., St John, K. and Gottesman, S. Regulation of RpoS by a novel small RNA: the characterization of RprA. *Mol. Microbiol*. (2001) 39, 1382–1394. doi.org/10.1111/j.1365-2958.2001.02329.x.

Majdalani, N., Cunning, C., Sledjeski, D., Elliott, T. and Gottesman, S. Dsra RNA regulates translation of RpoS message by an anti-antisense mechanism, independent

of its action as an antisilencer of transcription. *Proc. Natl. Acad. Sci. U.S.A.* (1998) 95: 12462–12467. doi.org/10.1073/pnas. 95.21.12462.

Mancuso, C., Pani, G., and Calabrese, V. Bilirubin: an endogenous scavenger of nitric oxide and reactive nitrogen species. *Redox Report* (2006) 11(5) doi: 10.1179/135100006X154978.

Mani, N., and Dupuy, B. Regulation of toxin synthesis in *Clostridium difficile* by an alternative RNA polymerase sigma factor. *P Natl Acad Sci USA* (2001) 98: 5844–9.

Mani, N., Lyras, D., Barroso, L., Howarth, P., Wilkins, T., Rood, J.I., Sonenshein, A.L. and Dupuy, B. Environmental response and autoregulation of *Clostridium difficile* TxeR, a sigma factor for toxin gene expression. *J Bacteriol* (2002) 184: 5971–8.

Marchais, A., Duperrier, S., Durand, S., Gautheret, D., and Stragier, P. CsfG, a sporulation-specific, small non-coding RNA highly conserved in endospore formers. *RNA Biology* (2011) 8(3): 358–364. doi:10.4161/rna.8.3.14998.

Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res*. (2015) 43(Database issue):D222–6.

Margolin, W. FtsZ and the division of prokaryotic cells and organelles. *Nat. Rev. Mol. Cell Biol.* (2005) 6:862–871. doi: 10.1038/nrm1745.

Martin-Verstraete, I., Peltier, J., and Dupuy, B. The regulatory networks that control *Clostridium difficile* toxin synthesis. *Toxins*. (2016) 8(5):pii:E153.

Martin, J., Monaghan, T. and Wilcox, M. *Clostridium difficile* infection: epidemiology, diagnosis and understanding transmission. *Nat Rev Gastroenterol Hepatol* (2016) 13, 206–216 doi.org/10.1038/nrgastro.2016.25.

Martin, W.F. and Sousa, F.L. Early Microbial Evolution: The Age of Anaerobes. Cold Spring *Harb Perspect Biol* (2016) 8:a018127.

Martinez-Melendez, A., Camacho-Ortiz, A., Morfin-Otero, R., Maldonado- Garza, H.J., Villarreal-Trevino, L., Garza-Gonzalez, E. Current knowledge on the laboratory diagnosis of *Clostridium difficile* infection. *World J Gastroenterol*. (2017)23: 1552–67.

Masse, E. and Gottesman, S. A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* (2002) 99(7): 4620–4625. doi.org/10.1073/pnas.032066599.

Maulucioni (2013) PhylogeneticTree, Woese 1990. Accessed May 2019. Available from:

Maziade, P. J., Pereira, P. and Goldstein, E. J. A decade of experience in primary prevention of *Clostridium difficile* infection at a community hospital using the probiotic combination *Lactobacillus acidophilus* CL1285, *Lactobacillus casei* LBC80R, and *Lactobacillus rhamnosus* CLR2 (Bio-K+). *Clin. Infect. Dis.* (2015) 60(2): S144–S147.

McBride, S.M., and Sonenshein, A.L. Identification of a genetic locus responsible for antimicrobial peptide resistance in *Clostridium difficile*. *Infect Immun*. (2011) 79(1): 167–76.

McCarty, J. S. and Walker, G. C. DnaK mutants defective in ATPase activity are defective in negative regulation of the heat shock response: expression of mutant DnaK proteins results in filamentation. *J. Bacteriol*. (1994) 176: 764-780.

McDonald, L.C., Coignard, B., Dubberke, E., Song, X., Horan, T., Kutty, P.K. and the *Ad Hoc Clostridium difficile* Surveillance Working Group. Recommendations for surveillance of *Clostridium difficile*-associated disease. *Infect Control Hosp Epidemiol*. (2007) 28(2):140-145. doi:10.1086/511798.

McDonald, L.C., Killgore, G.E., Thompson, A., Owens, R.C. Jr., Kazakova, S.V., Sambol, S.P., *et al*. An epidemic, toxin gene-variant strain of *Clostridium difficile*. *N Engl J Med.* (2005) 353: 2433-2441.

McKee, R.W., Harvest, C.K. and Tamayo, R. Cyclic Diguanylate Regulates Virulence Factor Genes via Multiple Riboswitches in *Clostridium difficile*. *mSphere.* (2018) 3:e00423-18. doi.org/10.1128/mSphere.00423-18.

Meader, E., Mayer, M.J., Gasson, M.J., Steverding, D., Carding, S.R., and Narbad, A. Bacteriophage treatment significantly reduces viable *Clostridium difficile* and prevents toxin production in an in vitro model system. *Anaerobe* (2010) 16: 549-54. doi.org/10.1016/j.anaerobe.2010.08.006.

Mellin, J.R., and Cossart, P. Unexpected versatility in bacterial riboswitches. *Trends Genet*. (2015) 31, 150–156. doi.org/10.1016/j.tig.2015.01.005.

Metcalf, D. and Weese, J.S. Evaluation of commercial kits for extraction of DNA and RNA from *Clostridium difficile. Anaerobe* (2012) 18(6):608-613. doi.org/10.1016/j.anaerobe.2012.10.003

Metcalf, D., Sharif, S., and Weese, J.S. (2010) Evaluation of candidate reference genes in *Clostridium difficile* for gene expression normalization. *Anaerobe*. 16: 439-443.

Michaux, C., Hartke, A., Martini, C., Reiss, S., Albrecht, D., Budin-Verneuil, A., Sanguinetti, M., Engelmann, S., Hain, T., Verneuil, N., Giardb, J., Involvement of *Enterococcus faecalis* Small RNAs in Stress Response and Virulence. *Infection and Immunity.* (2014) 82(9): 3599–3611.

Michaux, C., Verneuil, N., Hartke, A., and Giard, J. Physiological roles of small RNA molecules. *Microbiology* (2014) 160: 1007–1019. doi:10.1099/mic.0.076208-0.

Mohr, G., D. Smith, M. Belfort, and Lambowitz, A.M. Rules for DNA target-site recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. *Genes Dev.* (2000) 14: 559-73.

Molle, V., Fujita, M., Jensen, S.T., Eichenberger, P., González-Pastor, J.E., Liu, J.S., and Losick, R. The Spo0A regulon of *Bacillus subtilis*. *Mol. Microbiol*. (2003) 50: 1683-1701.

Monleón, D., Morales, J.M., Barrasa, A., López, J.A., Vázquez, C., and Celda, B. Metabolite profiling of fecal water extracts from human colorectal cancer. *NMR Biomed*. (2009) 22(3): 342-8

Monot M, Boursaux-Eude C, Thibonnier M, Vallenet D, Moszer I, Medigue C, *et al.* Reannotation of the genome sequence of *Clostridium difficile strain* 630. *J Med Microbiol*. (2011) 60(8):1193–9.

Moore, J.H., Pinheiro, C.C.D., Zaenker, E.I., Bolick, D.T., Kolling, G.L., van Opstal, E., *et al*. Defined nutrient diets alter susceptibility to *Clostridium difficile* associated disease in a murine model. *PLOS ONE*. (2015) 10(7): e0131829.

Mori, N., and Takahashi, T. Characteristics and Immunological Roles of Surface Layer Proteins in *Clostridium difficile*. *Ann Lab Med*. (2018) 38(3):189-195. doi: 10.3343/alm.2018.38.3.189.

Moszer I, Jones LM, Moreira S, Fabry C, and Danchin A. SubtiList: the reference database for the *Bacillus subtilis* genome. *Nucleic Acids Res*. 2002;30(1):62–5.

Mraheil, M.A., Billion, A., Kuenne, C., Pischimarov, J., Kreikemeyer, B., Engelmann, S., Hartke, A., Giard, C., Rupnik, M., Vorwerk, S., Beier, M., Retey, J., Hartsch, T., Jacob, A., Cemic, F., Hemberger, J., Chakraborty, T., and Hain, T. Comparative genome-wide analysis of small RNAs of major Gram positive pathogens: from identification to application. *Microbial Biotechnology* (2010) doi:10.1111/j.1751-7915.2010.00171.x

Mulhbacher, J., Brouillette, E., Allard, M., Fortier, L.C., Malouin, F., and Lafontaine, D.A. Novel riboswitch ligand analogs as selective inhibitors of guanine-related metabolic pathways. *PLoS Pathog* (2010) 6(4): e1000865. 10.1371/ journal.ppat.1000865.

Mullany, P., Allan, E. and Roberts, A.P. Mobile genetic elements in *Clostridium difficile* and their role in genome function. *Res Microbiol*. (2015) 166(4):361-367. doi:10.1016/j.resmic.2014.12.005.

Nagahama, M., Takehara, M., and Rood, J.I. Histotoxic clostridial infections. *Microbiol Spectrum* (2019) 6(4): GPP3-0024-2018. doi:10.1128/microbiolspec.GPP3-0024-2018.

Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L., and Breaker, R.R. Genetic control by a metabolite binding mRNA. *Chem Biol.* (2002) 9(9):1043. doi: 10.1016/s1074-5521(02)00224-7.

Napolitano, L. M. and Edmiston, C. E. *Clostridium difficile* disease: diagnosis, pathogenesis, and treatment update. *Surgery* (2017) 162: 325–348.

Neidhardt, F., C., Ingraham, J., L., and Schaechter, M. Physiology of the bacterial cell: A molecular approach. *Biochemistry and Molecular Biology Education*. (1992) 20:124-125.

NHS (2018) *Clostridium difficile.* Accessed 11th November 2020. Available from https://www.nhs.uk/conditions/c-difficile/

Northern Ireland Statistics and Research Agency (NISRA) (2018) 2008-2017 CDIFF Deaths. Accessed 5th February 2019. Available from: https://www.nisra.gov.uk/publications/2008-2017-cdiff-deaths

Nowak, A., Śliżewska, K., Błasiak, J., and Libudzisz, Z. The influence of *Lactobacillus casei* DN 114 001 on the activity of faecal enzymes and genotoxicity of faecal water in the presence of heterocyclic aromatic amines. *Anaerobe*. (2014) 30: 129-36.

Nuyts, S., Van Mellaerta, L., and Anne, J. Efficient isolation of total RNA from *Clostridium* without DNA contamination. *Journal of Microbiological Methods.* (2001) 44: 235–238.

O'Connor, J. R., Lyras, D., Farrow, K. A., Adams, V., Powell, D. R., Hinds, J., Cheung, J. K. and Rood, J. I. Construction and analysis of chromosomal *Clostridium difficile* mutants. *Mol Microbiol*. (2006) 61:1335-1351.

Pace, N.R. Mapping the Tree of Life: Progress and Prospects. *Microbiology and Molecular Biology Reviews* (2009) 73(4): 565–576. doi:10.1128/MMBR.00033-09.

Padalon-Brauch, G., Hershberg, R., Elgrably-Weiss, M., Baruch, K., Rosenshine, I., Margalit, H., and Altuvia, S. Small RNAs encoded within genetic islands of *Salmonella typhimurium* show host-induced expression and role in virulence. *Nucleic Acids Res.* (2008) 36**:** 1913-1927.

Paget, M.S. Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. *Biomolecules* (2015) 5(3):1245-65. doi:10.3390/biom5031245.

Pallen, M.J. and Matzke, N.J. From The Origin of Species to the origin of bacterial flagella. *Nat. Rev. Microbiol.* (2006) 4: 784-790.

Paredes-Sabja, D., Shen, A. and Sorg, J.A. *Clostridium difficile* spore biology: sporulation, germination, and spore structural proteins. *Trends Microbiol*. (2014) 22(7): 406–416. doi:10.1016/j.tim.2014.04.003.

Paredes, C.J., Alsaker, K.V., Papoutsakis, E.T.: A comparative genomic view of clostridial sporulation and physiology. *Nat Rev Microbiol* (2005) 3(12): 969-978.

Patenge, N., Billion, A., Raasch, P., Normann, J., Wisniewska-Kucper, A., Retey, J., Boisguérin, V., Hartsch, T., Hain, T., and Kreikemeyer, B. Identification of novel growth phase- and media- dependent small non-coding RNAs in Streptococcus pyogenes M49 using intergenic tiling arrays. BMC Genomics (2012) 13: 550.

Pavlostathis, S. G., Marchant, R., Banat, I., Ternan, N., and McMullan, G. High growth rate and substrate exhaustion results in rapid cell death and lysis in the thermophilic bacterium *Geobacillus thermoleovorans*. Biotechnology and Bioengineering (2006) 95(1): 84-95. doi.org/10.1002/bit.20962.

Pearson, J.R., Gill, C.I., and Rowland, I.R. Diet, fecal water, and colon cancer-development of a biomarker. *Nutr Rev*. (2009) 67(9): 509-26.

Peng, Z., Ling, L., Stratton, C.W., Li, C., Polage, C.R., Wu, B., and Tang, Y.W. Advances in the diagnosis and treatment of *Clostridium difficile* infections. *Emerging Microbes & Infections* (2018)7:15. doi:10.1038/s41426-017-0019-4

Pépin, J., Valiquette, L., Cossette, B. Mortality attributable to nosocomial *Clostridium difficile*-associated disease during an epidemic caused by a hypervirulent strain in Quebec. *Canadian Medical Association Journal*. (2005) 173:1037-1042.

Pereira, F. C., Saujet, L., Tome, A.R., Serrano, M., Monot, M., Couture-Tosi, E., Martin-Verstraete, I., Dupuy, B., and Henriques, A.O. The spore differentiation pathway in the enteric pathogen *Clostridium difficile*. *PLoS Genet*. (2013) 9: e1003782.

Permpoonpattana, P., Phetcharaburanin, J., Mikelsone, A., Dembek, M., Tan, S., Brisson, M.C., *et al*. Functional characterization of *Clostridium difficile* spore coat proteins. *J Bacteriol*. (2013)195(7):1492–503.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*. (2011) 8(10): 785-6.

Pettit, L.J., Browne, H.P., Yu, L., Smits, W.K., Fagan, R.P., Barquist, L., *et al*. Functional genomics reveals that *Clostridium difficile* Spo0A coordinates sporulation, virulence and metabolism. *BMC Genomics*. (2014) 15:160.

Pfeiffer, V., Papenfort, K., Lucchini, S., Hinton, J.C. and Vogel, J. Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat Struct Mol Biol*. (2009) 16: 840-846.

Pichon, C., and Felden, B. Small RNA genes expressed from *Staphylococcus aureus* genomic and pathogenicity islands with specific expression among pathogenic strains. *Proc Natl Acad Sci USA.* (2005) 102: 14249-14254.

Pishdadian, K., Fimlaid, K.A., and Shen, A. SpoIIID-mediated regulation of σK function during *Clostridium difficile* sporulation. *Mol Microbiol*. (2015) 95(2): 189-208.

Planche, T., Aghaizu, A., Holliman, R., and Riley, P. Diagnosis of *Clostridium difficile* infection by toxin detection kits: a systematic review. *Lancet Infect Dis.* (2008) 8: 777-784.

Pochapin, M. The effect of probiotics on *Clostridium difficile* diarrhea. *Am. J. Gastroenterol*. (2000) 95: S11–S13.

Popoff, M.R., Rubin, E.J., Gill, D.M., and Boquet, P. Actin-specific ADP-ribosyltransferase produced by a *Clostridium difficile* strain. *Infect Immun* (1988) 56: 2299306.

Popp, S., Packschies, L., Radzwill, N., Vogel, K., P., Steinhoff, H., J., Reinstein, J. Structural dynamics of the dnaK-peptide complex. *Journal of Molecular Biology* (2005) 347:1039-1052.

Poutanen, S. M. and Simor., A. E. *Clostridium difficile*-associated diarrhea in adults. *CMAJ* (2004) 171: 51-58.

Pozzoni, P., Riva, A,, Bellatorre, A.G., Amigoni, M., Redaelli, E., Ronchetti, A., Stefani, M., Tironi, R., Molteni, E.E., Conte, D., Casazza, G., and Colli, A. *Saccharomyces boulardii* for the prevention of antibiotic-associated diarrhea in adult hospitalized patients: a single-center, randomized, double-blind, placebo-controlled trial. *Am. J. Gastroenterol*. (2012) 107: 922–931.

Putral, L., Gu, W., and McMillan, N. RNA interference for the treatment of cancer. *Drug News Perspect.* (2006) 19 (6): 317–24. Doi:10.1358/dnp.2006.19.6.985937.

Putzer, H., Gendron, N., and Grunberg-Manago, M. Coordinate expression of the two threonyl-tRNA synthetase genes in *Bacillus subtilis*: control by transcriptional antitermination involving a conserved regulatory sequence. *EMBO J*. (1992) 11: 3117-3127.

Rafter, J.J., Child, P., Anderson, A.M., Alder, R., Eng, V., and Bruce, W.R. Cellular toxicity of fecal water depends on diet. *Am J Clin Nutr*. (1987) 45(3): 559–63.

Ramos-Silva, P., Serrano, M., Henriques, A.O. From root to tips: sporulation evolution and specialization in *Bacillus subtilis* and the intestinal pathogen *Clostridioides difficile. Molecular Biology and Evolution* (2019) doi.org/10.1101/473793.

Ramsay, I., Brown, N.M. and Enoch, D.A. Recent Progress for the Effective Prevention and Treatment of Recurrent *Clostridium difficile* Infection. *Infectious Diseases: Research and Treatment* (2018) 11: 1–4.

Ramsey, M., Dallal, M.D., Brian, G., Harbrecht, M.D., Arthur, J., Boujoukas, M.D., Carl, A., Sirio, M.D., Linda, M., Farkas, M.D., Kenneth, K., Lee, M.D., and Richard, L., and Simmons, M.D. Fulminant *Clostridium difficile*: An underappreciated and increasing cause of death and complications. *Ann Surg* (2002) 235: 363–372.

Rao, K., Erb-Downward, J.R., Walk, S.T., Micic, D., Falkowski, N., Santhosh, K., Mogle, J. A., Ring, C., Young, V. B., Huffnagle, G. B. and Aronoff, D. M. *et al*. The systemic inflammatory response to *Clostridium difficile* infection. *PLoS One*. (2014) 9(3):e92578. doi:10.1371/journal.pone.0092578.

Ratsep, M., Kõljalg, S., Sepp, E., Smidt, I., Truusalu, K., Songisepp, E., Stsepetova, J., Naaber, P., Mikelsaar, R.H., Mikelsaar, M. A combination of the probiotic and prebiotic product can prevent the germination of *Clostridium difficile* spores and infection. *Anaerobe* (2017) 47: 94–103.

Rau, M.H., Bojanovič, K., Toftgaard Nielsen, A., and Long, K.S. Differential expression of small RNAs under chemical stress and fed-batch fermentation in *E. coli.* *BMC Genomics* (2015) 16:1051. DOI 10.1186/s12864-015-2231-8.

Razaq, N., Sambol, S., Nagaro, K., Zukowski, W., Cheknis, A., Johnson, S., *et al.* Infection of hamsters with historical and epidemic BI types of *Clostridium difficile*. *J Infect Dis.* (2007) 196: 1813-9.

Repoila, F., Majdalani, N., and Gottesman, S. Small non-coding RNAs, co-ordinators of adaptation processes in *Escherichia coli*: the RpoS paradigm. *Mol. Microbiol.* (2003) 48 (4): 855–861.

Rifkin, G.D., Silva, J. Jr., and Fekety. R. Gastrointestinal and systemic toxicity of fecal extracts from hamsters with clindamycin-induced colitis. *Gastroenterology* (1978) 74: 52-7.

Rosch, J.W., Vega, L.A., Beyer, J.M., Lin, A., and Caparon, M. G. The Signal Recognition Particle Pathway Is Required for Virulence in *Streptococcus pyogenes.* *Infection and Immunity* (2008) 76(6): 2612–2619.

Rosenbusch, K.E., Bakker, D., Kuijper, E.J., Smits, W.K. *C. difficile* 630Deltaerm Spo0A regulates sporulation, but does not contribute to toxin production, by direct high-affinity binding to target DNA. PloS One (2012) 7(10): e48608.

Rossen, J.W.A., Friedrich, A.W. and Moran-Gilad, J. Practical issues in implementing whole-genome-sequencing in routine diagnostic microbiology. *Clinical Microbiology and Infection* (2018) 24(4) P355-360. doi.org/10.1016/j.cmi.2017.11.001.

Rupnik, M., Brazier, J.S., Duerden, B.I., Grabnar, M., and Stubbs, S. Comparison of toxinotyping and PCR ribotyping of Clostridium difficile strains and description of novel toxinotypes. *Microbiology*. (2001) 147: 439-447.

Rupnik, M., Wilcox, M.H., and Gerding, D.N. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat Rev Microbiol*. (2009) 7: 526-36.

Sabat, A.J., van Zanten, E., Akkerboom, V., Wisselink, G., van Slochteren, K., de Boer, R.F. *et al.* Targeted next-generation sequencing of the 16S-23S rRNA region for

culture-independent bacterial identification - increased discrimination of closely related species. *Sci Rep.* (2017) 7(3434–017-03458-6).

Sadeghifard, N., Gürtler, V., Beer, M., and Seviour, R.J. The mosaic nature of intergenic 16S-23S rRNA spacer regions suggests rRNA operon copy number variation in *Clostridium difficile* strains. *Appl. Environ. Microbiol*. (2006) 72:7311–7323.

Samarkos, M., Mastrogianni, E., and Kampouropoulou, O. The role of gut microbiota in *Clostridium difficile* infection. *European Journal of Internal Medicine.* (2018) 50: 28-32.

Sambol, S.P., Tang, J.K., Merrigan, M.M., Johnson, S., and Gerding, D.N. Infection of hamsters with epidemiologically important strains of *Clostridium difficile*. *J Infect Dis*. (2001) 183: 1760-6. doi.org/10.1086/320736

Sandhu, B.K., and McBride, S.M. *Clostridioides difficile. Trends in Microbiology* (2018) 26(12): 1049-1050.

Sartelli, M., Malangoni, M.A., Abu-Zidan, F.M., Griffiths, E.A., Di Bella, S. *et al*. WSES guidelines for management of *Clostridium difficile* infection in surgical

patients. *World Journal of Emergency Surgery* (2015) 10:38 doi 10.1186/s13017-015-0033-6.

Saujet, L., Pereira, F.C., Henriques, A.O., and Martin-Verstraete, I. The regulatory network controlling spore formation in *Clostridium difficile*. *FEMS Microbiol Lett*. (2014) 358(1):1–10.

Schonfeld, H., Schmidt, D., Schroder, H. and Bukau, B. The DnaK Chaperone System of *Escherichia coli*: Quaternary Structures and Interactions of the DnaK and GrpE Components. *The Journal of Biological Chemistry* (1995) 270: 2183-2189. doi: 10.1074/jbc.270.5.2183.

Schulz, A., Tzschaschel, B. and Schumann, W. Isolation and analysis of mutants of the *dnaK* operon of *Bacillus subtilis*. *Mol. Microbiol*. (1995) 15: 421-429.

Sebaihia, M., Peck, M.W., Minton, N.P., Thomson, N.R., Holden, M.T., *et al.* Genome sequence of a proteolytic (Group 1) *Clostridium botulinum* strain Hall A and comparative analysis of the clostridial genomes. *Genome Res.* (2007) 17: 1082-1092.

Sebaihia, M., Wren, B.W., Mullany, P., Fairweather, N.F., Minton, N.P., Stabler, R., *et al.* The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat Genet.* (2006) 38:779–86.

Sedlyarova, N., Shamovsky, I., Bharati, B.K., Epshtein, V., Chen, J., Gottesman, S., Schroeder, R., and Nudler, E. sRNA-Mediated Control of Transcription Termination in *E. coli. Cell* (2016) 167, 111–121. doi.org/10.1016/j.cell.2016.09.004.

Selby, K., Lindstrom, M., Somervuo, P., Heap, J.T., Minton, N.P., and Korkeala, H. Important role of class I heat shock genes *hrcA* and *dnaK* in the heat shock response and the response to pH and NaCl stress of group I *Clostridium botulinum* strain ATCC 3502. *Appl. Environ. Microbiol*. (2011) 77: 2823–2830.

Setten, R.L., Rossi, J.J. and Han, S. The current state and future directions of RNAi-based therapeutics. *Nat Rev Drug Discov.* (2019). 18: 421–446 doi.org/10.1038/s41573-019-0017-4

Shi, W., Zhou, Y., Wild, J., Adler, J. and Gross, C. A. DnaK, DnaJ, and GrpE are required for flagellum synthesis in *Escherichia coli*. *J. Bacteriol*. (1992) 174: 6256-6263.

Shimizu, T., Ohtani, K., Hirakawa, H., Ohshima, K., Yamashita, A., Shiba, T., Ogasawara, N., Hattori, M., Kuhara, S., and Hayashi, H. Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Proc Natl Acad Sci USA*. (2002) 99: 996-1001.

Shimizu, T., Yaguchi, H., Ohtani, K., Banu, S., and Hayashi, H. Clostridial VirR/VirS regulon involves a regulatory RNA molecule for expression of toxins. *Mol. Microbiol.* (2002) 43**:** 257–265.

Singh, M., Vaishnavi, C., Mahmood, S. and Kochhar, R. Toxinotyping and Sequencing of *Clostridium difficile* Isolates from Patients in a Tertiary Care Hospital of Northern India. *Front Med (Lausanne)* (2017) 4: 33. doi: 10.3389/fmed.2017.00033.

Smits, W. K., Lyras, D., Lacy, D. B., Wilcox, M. H. and Kuijper, E. J. *Clostridium difficile* infection. *Nat. Rev. Dis. Primers* (2016) 2: 16020.

Song, Y. PCR-based diagnostics for anaerobic infections. *Anaerobe* (2005) 11:79-91.

Sorg, A., Sonenshein, A.L. Bile Salts and Glycine as Cogerminants for *Clostridium difficile* spores. *J Bacteriol*. (2008) 190(7): 2501-2512.

Soto, S. M. Role of efflux pumps in the antibiotic resistance of bacteria embedded in a biofilm. *Virulence* (2013) 4(3), 223–229. doi.org/10.4161/viru.23724.

Soutourina, O.A., Monot, M., Boudry, P., Saujet, L., Pichon, C., Sismeiro, O., Semenova, E., Severinov, K., Le Bouguenec, C., Coppee, J., Dupuy, B., Martin-Verstraete, I. Genome-Wide Identification of Regulatory RNAs in the Human Pathogen *Clostridium difficile. PLOS Genetics* (2013) 9 (5): e1003493.

Stabler, R. A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., Lawley, T.D., Sebaihia, M., Quail, M.A., Rose, G., Gerding, D.N., Gibert, M., Popoff, M.R., Parkhill, J., Dougan, G., and Wren, B.W. Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *BMC Genome Biol.* (2009) 10: R102. doi:10.1186/gb-2009-10-9-r102.

Stabler, R.A., Gerding, D.N., Songer, J.G., Drudy, D., Brazier, J.S., Trinh, H.T., Witney, A.A., Hinds, J., and Wren, B.W. Comparative phylogenomics of *Clostridium difficile* reveals clade specificity and microevolution of hypervirulent strains. *J Bacteriol*. (2006) 188: 7297-7305.

Stamatopoulou, V., Apostolidi, M., Li, S., Lamprinou, K., Papakyriakou, A., Zhang, J., and Stathopoulos, C. Direct modulation of T-box riboswitch-controlled transcription by protein synthesis inhibitors. *Nucleic Acids Research* (2017) 45 (17) 10242–10258. doi:10.1093/nar/gkx663.

Stark, R., Grzelak, M. and Hadfield, R. RNA sequencing: the teenage years. *Nature Reviews*. (2019) 20: 631-656.

Steiner, E., Dago, A.E., Young, D.I., Heap, J.T., Minton, N.P., Hoch, J.A. and Young, M. Multiple orphan histidine kinases interact directly with Spo0A to control the initiation of endospore formation in *Clostridium acetobutylicum*. *Mol Microbiol.* (2011) 80: 641– 654.

Storz, G., Opdyke, J. A. and Zhang, A. Controlling mRNA stability and translation with small, noncoding RNAs. *Curr. Opin. Microbiol*. (2004) 7: 140-144.

Suárez, J.M., Edwards, A.N., and McBride, S.M. The *Clostridium difficile cpr* locus is regulated by a noncontiguous two–component system in response to type A and B lantibiotics. *J Bacteriol*. (2013) 195(11): 2621–31.

Sugimoto, S., Saruwatari, K., Higashi, C. and Sonomoto, K. The proper ratio of GrpE to DnaK is important for protein quality control by the DnaK-DnaJ-GrpE chaperone system and for cell division. *Microbiology.* (2008) 154: 1876-1885.

Surawicz, C.M., Brandt, L.J., Binion, D.G., Ananthakrishnan, A.N., Curry, S.R., Gilligan, P.H., McFarland, L.V., Mellow, M., Zuckerbraun, B.S. Guidelines for

diagnosis, treatment, and prevention of *Clostridium difficile* infections. *Am. J. Gastroenterol*. (2013) 108, 478–498.

Tacconelli, E., and Cataldo, M. Vancomycin-resistant enterococci (VRE): transmission and control. *International Journal of Antimicrobial Agents*. (2008) 31(2): 99- 106.

Tan, K.S., Wee, B.Y., and Song, K.P. Evidence for holin function of *tcdE* gene in the pathogenicity of *Clostridium difficile*. *J Med Microbiol*. (2001) 50: 613-619.

Tasteyre, A. Barc, M.C., Karjalainen, T., Dodson, P., Hyde, S., Bourlioux, P., and Borriello, P. A. A Clostridium difficile gene encoding flagellin. *Microbiology* (2000) 146: 957-966.

Tasteyre, A., Barc, M. C., Collignon, A., Boureau, H. and Karjalainen, T. Role of FliC and FliD flagellar proteins of *Clostridium difficile* in adherence and gut colonization. *Infect. Immun*. (2001) 69: 7937-7940.

Ternan, N.G., Jain, S., Graham, R.L.J., and McMullan, G. Semiquantitative analysis of clinical heat stress in *Clostridium difficile* strain 630 using a GeLC/MS workflow with emPAI quantitation. *PLOS ONE*. (2014) 9(2):e88960.

Ternan, N.G., Jain, S., Srivastava, M., and McMullan, G. Comparative transcriptional analysis of clinically relevant heat stress response in *Clostridium difficile* Strain 630. *PLOS ONE*. (2012) 7(7): e42410.

Ternan, N.G., Moore, N.D., Smyth, D., McDougall, G.J., Allwood, J.W., Verrall, S., Gill, C.I.R., Dooley, J.S.G. and McMullan, G. Increased sporulation underpins adaptation of *Clostridium difficile* strain 630 to a biologically–relevant faecal environment, with implications for pathogenicity. *Scientific Reports* (2018) 8: 16691 doi:10.1038/s41598-018-35050-x.

Ternan. N.G., Jain, S., Srivastava, M., and McMullan, G. Comparative transcriptional analysis of clinically relevant heat stress response in *Clostridium difficile* strain 630. *PLOS ONE*. (2012) 7(7):e42410.

Thomason, M.K. and Storz, G. Bacterial antisense RNAs: How many are there and what are they doing? *Annu. Rev. Genet*. (2010) 44: 167–188. doi:10.1146/annurev-genet-102209-163523.

Tillotson, G.S., and Tillotson, J. *Clostridium difficile* - a moving target. *Med Rep* (2011) 3: 6.

Trevino, J., Perez, N., Sumby, P. The 4.5S RNA component of the signal recognition particle is required for group A Streptococcus virulence. Microbiology (2010) 156:1342–1350.

Twine, S.M., Reid, C.W., Aubry, A., McMullin, D.R., Fulton, K.M., Austin, J., and Logan, S.M. Motility and flagellar glycosylation in *Clostridium difficile. J. Bacteriol.* (2009) 191: 7050-7062.

Underwood, S., Guan, S., Vijayasubhash, V., Baines, S.D., Graham, L., Lewis, R.J., Wilcox, M.H., and Stephenson, K. Characterization of the sporulation initiation pathway of *Clostridium difficile* and its role in toxin production. *Journal of bacteriology*. (2009) 191: 7296-7305.

Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R. and Leunissen, J.A.M. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res*. (2007) 35. W71-W74.

Valencia-Sanchez, M., Liu, J., Hannon, G., and Parker, R. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.* (2006) 20: 515-524.

Vedantam, G., Clark, A., Chu, M., McQuade, R., Mallozzi, M., and Viswanathan, V. K. Clostridium Difficile infection: toxins and non-toxin virulence factors, and their

contributions to disease establishment and host response. *Gut Microbes* (2012) 3: 121–134. doi:10.4161/gmic.19399.

Verbeke, K.A., Boobis, A.R., Chiodini, A., Edwards, C.A., Franck, A., Kleerebezem, M, *et al*. Towards microbial fermentation metabolites as markers for health benefits of prebiotics. *Nutr Res Rev*. (2015) 28(1): 42-66.

Vermeiren, J., Van de Wiele, T., Verstraete, W., Boeckx, P., and Boon, N. Nitric Oxide Production by the Human Intestinal Microbiota by Dissimilatory Nitrate Reduction to Ammonium. Journal of Biomedicine and Biotechnology (2009) doi:10.1155/2009/284718.

Villano, S.A., Seiberling, M., Tatarowicz, W., Monnot-Chase, E. and Gerding, D.N. Evaluation of an oral suspension of VP20621, spores of nontoxigenic *Clostridium difficile* strain M3, in healthy subjects. *Antimicrobial Agents Chemother*. (2012) 10:5224–5229.

Vogel, J. A rough guide to the non-coding RNA world of Salmonella. *Mol. Microbiol.* (2009) 71 (1): 1-11.

Vogel, J., and Wagner, E.G. Target identification of small noncoding RNAs in bacteria. *Curr. Opin. Microbiol*. (2007) 10 (3): 262–270.

Voorhoeve, P.M., and Agami, R. Knockdown stands up. *Trends Biotechnol.* (2003) 21 (1): 2–4.

Wakefield, A.J, Murch, S.H, Anthony, A., Linnell, .J, Casson, D.M, Malik, M., Berelowitz, M., Dhillon, A.P., Thomson, M.A., Harvey, P., Valentine, A., Davies, S.E. and Walker-Smith, J.A. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet* (1998) 28: 637-41.

Wan, Y., Kertesz, M., Spitale, R.C., Segal, E., and Chang, H.W. Understanding the transcriptome through RNA structure. *Nature.* (2011) 12: 641-655.

Warny, M., Pepin, J., Fang, A., Kilgore, G., Thompson, A., Brazier, J., Frost, E., and McDonald, L.C. Toxin production by an emerging strain of *Clostridium difficile* associated with outbreaks of severe disease in North America and Europe. *Lancet* (2005) 366:1079-1084.

Waters, L. S., and G. Storz. Regulatory RNAs in bacteria. *Cell*. (2009) 136: 615– 628.

Weiss, M.C., Sousa, F.L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-Sathi, S., and Martin, W.F. The physiology and habitat of the last universal common ancestor. *Nature Microbiology* (2016) 1 (9): 16116. doi:10.1038/nmicrobiol.2016.116.

Wilcox, M.H. Updated guidance on the management and treatment of Clostridium difficile infection. Public Health England. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/321891/Clostridium_difficile_management_and_treatment.pdf Accessed September 2019.

Wilcox, M.H., Ahir, H., Coia, J.E., Dodgson, A., Hopkins, S., Llewelyn, M.J., Settle, C., Mclain-Smith, S., and Marcella, S.W. Impact of recurrent *Clostridium difficile* infection: hospitalization and patient quality of life. *Journal of Antimicrobial Chemotherapy*. (2017) 72(9): 2647–2656 doi.org/10.1093/jac/dkx174.

Williamson, G., and Clifford, M.N. Role of the small intestine, colon and microbiota in determining the metabolic fate of polyphenols. *Biochem Pharmacol*. (2017) 139:24-39.

Windey, K., De Preter, V., Huys, G., Broekaer,t WF., Delcour, J.A., Louat, T., *et al*. Wheat bran extract alters colonic fermentation and microbial composition, but does not affect faecal water toxicity: a randomised controlled trial in healthy subjects. *Br J Nutr*. (2015) 113(2): 225-38.

Woese, C.R. Bacterial evolution. *Microbiol Rev* (1987) 51: 221-271.

Woese, C.R., and Fox, G.E. Phylogenetic structure of prokaryotic domain - Primary kingdoms. *Proc Natl Acad Sci* (1977) 74: 5088–5090.

Woese, C.R., Kandler, O., and Wheelis, M.L. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Nati. Acad. Sci. USA* (1990) 87: 4576-4579.

Wolf, D., Bröning, T., and Gerritzen, A. Rapid detection of the *Clostridium difficile* ribotype 027 tcdC gene frame shift mutation at position 117 by real-time PCR and melt curve analysis. *Eur J Clin Microbiol Infect Dis*. (2009) 28: 959-962.

Wullt, M., Hagslatt, M. L. and Odenholt, I. *Lactobacillus plantarum* 299v for the treatment of recurrent *Clostridium difficile*-associated diarrhoea: a double-blind, placebo-controlled trial. *Scand. J. Infect. Dis.* (2003) 35: 365–367.

Yamane, K., Bunai, K. and Kakeshita, H. Protein traffic for secretion and related machinery of *Bacillus subtilis*. *Biosci. Biotechnol. Biochem*. (2004) 68: 2007–2023.

Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R., *et al*. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics*. (2010) 26(13): 1608-15.

Yutin, N. and Galperin, M.Y. A genomic update on clostridial phylogeny: Gram-negative spore formers and other misplaced clostridia. *Environ Microbiol*. (2013) 15(10):2631-2641. doi:10.1111/1462-2920.12173.

Zhang, J., and Ferré-D'Amaré, A.R. Structure and mechanism of the T-box riboswitches. *Wiley Interdiscip Rev RNA* (2015) 6(4): 419–433. doi:10.1002/wrna.1285.

Zhong, J., M. Karberg, and Lambowitz, A.M. Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Res*. (2003) 31:1656-64.

Zhu, D., Sorg, J.A., and Sun, X. *Clostridioides difficile* Biology: Sporulation, Germination, and Corresponding Therapies for *C. difficile* Infection. *Front. Cell. Infect. Microbiol*. (2018) 8(29):1-10. doi.org/10.3389/fcimb.2018.00029.

Ziebandt, A.K., Becher, D., Ohlsen, K., Hacker, J., Hecker, M., and Engelmann, S. The influence of *agr* and sigmaB in growth phase dependent regulation of virulence factors in *Staphylococcus aureus*. *Proteomics* (2004) 4: 3034-3047.